

LIRIS

UMR 5205 CNRS

Thèses et HDR 2017

Laboratoire d'InfoRmatique
en Image et Systèmes d'information



INSA



UNIVERSITÉ
LUMIÈRE
LYON 2



HDR

Contributions to Building Reliable Distributed Systems Sonia Ben Mokhtar	5
Empowering Ambivalence - Supporting multiple interpretations in knowledge-based systems Pierre-Antoine Champin	9
Modélisation de terrains virtuels Eric Guérin	17

Thèses

Méthodologie de développement de logiciels dans un environnement informatiquement vert Hayri Acar	21
Une approche générique pour la construction collaborative d'éditions numériques critiques Vincent Barreton	25
Decomposition and Domination of Some Graphs Fairouz Beggas	29
Intégration des données hétérogènes issues de plusieurs fournisseurs de services géo-localisés : un cas d'utilisation sur les points d'intérêt touristique Bilal Berjawi	31
Anytime discovery of a diverse set of patterns with Monte Carlo tree search Guillaume Bosc	35
Compression progressive de maillages surfaciques texturés Florian Caillaud	39
Mécanismes de négociation distribuée pour la Gestion intelligente de l'énergie Romain Cailliere	43

Détection d'objets en mouvement à l'aide d'une caméra mobile Marie-Neige Chapel	45
Approche déclarative pour le stockage à long terme de données capteur Manel Charfi	47
Acquisition de connaissances pour des EIAH destinés à enseigner des méthodes de résolution de problèmes Awa Diattara	51
Estimation de pose humaine et reconnaissance d'actions par un système multi-robots Emre Dogan	55
Traitement des graphes massifs : partitionnement et requêtage agrégatif Ghizlane Echbarthi	57
Apprentissage de structure de modèles graphiques probabilistes : application à la classification multi-label Maxime Gasse	59
Apprentissage multi-label ensembliste dans le contexte supervisé et semi-supervisé Ouadie Gharroudi	61
Traitement joint de nuages de points et d'images pour l'analyse et la visualisation de formes 3D Maximilien Guislain	63
Analyse automatisée de la circulation automobile par caméras de surveillance routière Kannikar Intawong	65
Addressing Selfishness in the Design of Cooperative Systems Guido Lena Cota	67
Conception et développement d'interactions immersives pour jeux sérieux Guillaume Loup	69
Transfert de connaissances pour la classification des images Ying Lu	71
Une approche orientée service pour la recherche sémantique de contenus multimédias Sid Ahmed Djalal Midouni	75

Utilisation de méthodes de décomposition pour les problèmes du plus grand sous-graphe commun et de la somme coloration	
Maël Minot	77
Isomorphisme de sous-graphes dans les graphes de données massifs	
Chemseddine Nabti	81
Représentations et dynamique de la ville virtuelle	
Frédéric Pedrinis	83
Subspace clustering on static datasets and dynamic data streams using bio-inspired algorithms	
Sergio Peignier	85
Introducing Privacy in Current Web Search Engines	
Albin Petit	89
Single image super-resolution based on neural networks for text and face recognition	
Clément Peyrard	91
Architecture de monitoring déclaratif durable : optimisation énergétique des interactions entre requêtes applicatives orientées service et réseau de capteurs sans fil - Application aux bâtiments intelligents	
Ozgun Pinarer	93
Étude de l'évolution des micro-organismes bactériens par des approches de modélisation et de simulation informatique	
Charles Rocabert	95
Adaptation contextuelle multi-préoccupations orientée sémantique dans le web des objets	
Mehdi Terdjimi	97
Modèles et outils pour favoriser l'articulation entre la généricité d'un assistant intelligent et les spécificités de son usage en EIAH	
Le Vinh Thai	101
Conception et analyse d'activités collaboratives pour environnement multi-surfaces	
Lili Tong	103
Capitalisation pérenne de connaissances industrielles : vers des méthodes de conception incrémentales et itératives centrées sur l'utilisateur	
Carine Touré	105

Correction de biais de sélection par pondération d'importance en apprentissage supervisé	
Van-Tinh Tran	107
Catégorisation des comportements de conduite en termes de consommation en carburant : une méthode de découverte de connaissances contextuelles à partir des traces d'interactions	
Assitan Traore	109
Copier-Coller 3D : paramétrisation cohérente de maillages triangulaires	
Ana-Maria Vintescu	111
Apprentissage et lissage dans les modèles de Markov à sauts et à copules	
Fei Zheng	113

Contributions to Building Reliable Distributed Systems

Sonia Ben Mokhtar

Jury

Rachid GUERRAOUI, PR, EPFL Lausanne
Karama KANOUN, Docteur, LAAS
Paulo ESTEVES-VERISSIMO, PR, Université du Luxembourg
Gordon BLAIR, PR, Lancaster University
Lionel BRUNIE, PR, INSA Lyon
Pascal FELBER, PR, Université de Neuchâtel
Valérie ISSARNY, Docteur, INRIA Silicon Valley
Gilles MULLER, Docteur, INRIA Paris
Quéma VIVIEN, PR, Grenoble INP

Rapporteur
Rapporteur
Rapporteur
Examineur
Examineur
Examineur
Examinatrice
Examineur
Examineur

Abstract

Thanks to the latest evolutions in hardware and networking technologies we are living in a world where networked computing systems are everywhere ranging from small/medium daily objects (e.g., watches, smart phones, cars) to large infrastructures (e.g., cloud platforms and data centers). On top of these computing systems a plethora of software systems/applications are invading our daily lives. Because of their intrinsic distribution and the involvement of more and more parties with sometimes conflicting interests, these systems are becoming bigger and increasingly more complex and thus more subject to faults. Beyond crash faults, two types of faults are often distinguished: Byzantine faults and rational faults. Byzantine faults are the most generic type of faults caused by nodes (e.g., software or hardware components running in a physical machine participating in the system) that may behave arbitrarily (e.g., by crashing, being subject to a bug, being under the control of a malicious attacker). Rational faults are caused by nodes trying to maximize their own benefit without contributing their fair share to the system. Dealing with Byzantine and/or rational faults in large scale distributed systems has been and still is a very active field of research. In the context of this habilitation defense I will present a set of research works I carried out since I joined the LIRIS lab with my PhD students and collaborators and that contribute to this field. I will finally sketch few perspectives in this very challenging and exciting research domain.

Publications

Revue internationale avec comité de lecture

- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar, Lionel Brunie & Ammar Hasan (2016). « 4PR: Privacy Preserving Routing in Mobile Delay Tolerant Networks ». *Computer Networks* (Elsevier). doi : 10.1016/j.comnet.2016.08.005. HAL : hal-01535796.
- Albin Petit, Thomas Cerqueus, Antoine Boutet, Sonia Ben Mokhtar, David Coquil, Lionel Brunie & Harald Kosch (2016). « SimAttack: private web search under fire ». *Journal of Internet Services and Applications*, p. 17. doi : 10.1186/s13174-016-0044-x. HAL : hal-01304320.
- Hayam Mousa, Sonia Ben Mokhtar, Omar Hasan, Osama Younes, Hadhoud Mohiy & Lionel Brunie (2015). « Trust management and reputation systems in mobile participatory sensing applications: A survey ». *Elsevier Computer Networks*, vol. 90, pp. 49-73. HAL : hal-01265403.
- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar, Lionel Brunie & Gabriele Gianini (2014). « A Delay and Cost Balancing Protocol for Message Routing in Mobile Delay Tolerant Networks ». *Ad Hoc Networks* (Elsevier), pp. 430-443. HAL : hal-01301078.

HDR

Soutenu le 21/12/2017

Établissement :
INSA Lyon

Contact :

Sonia.Benmokhtar@liris.cnrs.fr

- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar, Lionel Brunie & Kangbin Yim (2013). « An investigation on the unwillingness of nodes to participate in mobile delay tolerant network routing ». *International Journal of Information Management* (Elsevier), vol. 33, pp. 252-262. doi : 10.1016/j.ijinfomgt.2012.11.001. HAL : hal-01351700.
- Afra Mashhadi, Sonia Ben Mokhtar & Licia Capra (2012). « Fair Content Dissemination in Participatory DTNs ». *Elsevier Ad-Hoc Networks Journal*, vol. 10, pp. 1633-1645. HAL : hal-01352964.
- Sonia Ben Mokhtar, Pierre-Guillaume Raverdy, Aitor Urbieto & Roberto Speicys Cardoso (2010). « Interoperable Semantic and Syntactic Service Discovery for Ambient Computing Environments ». *International Journal of Ambient Computing and Intelligence*, vol. 2, pp. 13-32. HAL : hal-01381593.

Conférences internationales avec comité de lecture

- Sophie Cerf, Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, Sara Bouchenak, Nicolas Marchand & Bogdan Robu (2017). « Données de mobilité : protection de la vie privée vs. utilité des données ». Conférence francophone d'informatique en parallélisme, architecture et système (CompAS), Sophia Antipolis (France). HAL : hal-01527666.
- Sophie Cerf, Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar, Robert Birke, Sara Bouchenak, Lydia Chen, Nicolas Marchand & Bogdan Robu (2017). « PULP: Achieving Privacy and Utility Trade-off in User Mobility Data ». *SRDS 2017 - 36th IEEE International Symposium on Reliable Distributed Systems*, 29 septembre 2017, Hong Kong (Hong-Kong). HAL : hal-01578635.
- Guido Cota, Sonia Ben Mokhtar, Gabriele Gianini, Ernesto Damiani, Julia Lawall, Gilles Muller & Lionel Brunie (2017). « Analysing Selfishness Flooding with SEINE ». *The 47th IEEE/IFIP International Conference on Dependable Systems and Networks (DSN'17)*, 30 juin 2017, Denver, Colorado (États-Unis), pp. 603-614. doi : 10.1109/DSN.2017.51. HAL : hal-01581628.
- Sonia Ben Mokhtar, Antoine Boutet, Louafi Bouzouina, Patrick Bonnel, Olivier Brette, Lionel Brunie, Mathieu Cunche, Stephane D 'alu, Vincent Primault, Patrice Raveneau et al. (2017). « PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets ». *NetMob 2017*, 7 avril 2017, Milan (Italie). HAL : hal-01578557.
- Hayam Mousa, Sonia Ben Mokhtar, Omar Hasan, Lionel Brunie, Osama Shible Youness & Hadhoud Mohiy (2017). « A Reputation System Resilient Against Colluding and Malicious Adversaries in Mobile Participatory Sensing Applications ». *The 14th Annual IEEE Consumer Communications & Networking Conference (CCNC 2017)*, 8 janvier 2017, Las Vegas (États-Unis). HAL : hal-01535819.
- Vincent Primault, Antoine Boutet, Sonia Ben Mokhtar & Lionel Brunie (2016). « Adaptive Location Privacy with ALP ». *35th Symposium on Reliable Distributed Systems*, 29 septembre 2016, Budapest (Hongrie). ArXiv : 1609.07349. HAL : hal-01370447.
- Antoine Boutet, Albin Petit, Sonia Ben Mokhtar & Léa Laporte (2016). « Leveraging Query Sensitivity for Practical Private Web Search ». *Middleware*, 16 décembre 2016, Trento (Italie). doi : 10.1145/3007592.3007595. HAL : hal-01381995.
- Jérémie Decouchant, Sonia Ben Mokhtar, Albin Petit & Vivien Quéma (2016). « PAG: Private and Accountable Gossip ». *ICDCS - International Conference on Distributed Computing Systems*, 30 juin 2016, Nara (Japon). doi : 10.1109/ICDCS.2016.34. HAL : hal-01368911.

- Sophie Cerf, Bogdan Robu, Nicolas Marchand, Antoine Boutet, Vincent Primault, Sonia Ben Mokhtar & Sara Bouchenak (2016). « Toward an Easy Configuration of Location Privacy Protection Mechanisms ». ACM/IFIP/USENIX Middleware conference, 16 décembre 2016, Trente (Italie). HAL : hal-01376640.
- Guido Lena Cota, Sonia Ben Mokhtar, Julia Lawall, Gilles Muller, Gabriele Gianini, Ernesto Damiani & Lionel Brunie (2015). « A Framework for the Design Configuration of Accountable Selfish-Resilient Peer-to-Peer Systems ». SRDS 2015 - 34th International Symposium on Reliable Distributed Systems, 1 octobre 2015, Montreal (Canada). HAL : hal-01250717.
- Vincent Primault, Sonia Ben Mokhtar, Cédric Lauradoux & Lionel Brunie (2015). « Time Distortion Anonymization for the Publication of Mobility Data with High Utility ». 14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications, 22 août 2015, Helsinki (Finlande). ArXiv : 1507.00443. HAL : hal-01170060.
- Vincent Primault, Sonia Ben Mokhtar & Lionel Brunie (2015). « Privacy-preserving Publication of Mobility Data with High Utility ». 2015 35th IEEE International Conference on Distributed Computed Systems, 2 juillet 2015, Columbus, Ohio (États-Unis), pp. 802-803. ArXiv : 1506.09074. HAL : hal-01159291.
- Vincent Primault, Sonia Ben Mokhtar, Cédric Lauradoux & Lionel Brunie (2014). « Differentially Private Location Privacy in Practice ». Third Workshop on Mobile Security Technologies (MoST) 2014, 17 mai 2014, San Jose (États-Unis). ArXiv : 1410.7744. HAL : hal-01148230.
- Sonia Ben Mokhtar, Jérémie Decouchant & Vivien Quéma (2014). « AcTinG: Accurate Freerider Tracking in Gossip ». Symposium on Reliable Distributed Systems (SRDS 2014), 6 octobre 2014, Nara (Japon). doi : 10.1109/SRDS.2014.12. HAL : hal-01198451.
- Nicolas Haderer, Vincent Primault, Patrice Raveneau, Christophe Ribeiro, Romain Rouvoy & Sonia Ben Mokhtar (2014). « Towards a Practical Deployment of Privacy-preserving Crowd-sensing Tasks ». Middleware Posters and Demos '14, 8 décembre 2014, Bordeaux (France). doi : 10.1145/2678508.2678530. HAL : hal-01095787.
- Albin Petit, Sonia Ben Mokhtar, Lionel Brunie & Harald Kosch (2014). « Towards Efficient and Accurate Privacy Preserving Web Search ». Proceedings of the 9th Workshop on Middleware for Next Generation Internet Computing, 8 décembre 2014, Bordeaux, France (France), 1:1-1:6. doi : 10.1145/2676733.2676734. HAL : hal-01301099.
- Amadou Diarra, Sonia Ben Mokhtar, Pierre-Louis Aublin & Vivien Quema (2014). « FullReview: Practical Accountability in Presence of Selfish Nodes ». The 33rd IEEE Symposium on Reliable Distributed Systems (SRDS 2014), 6 octobre 2014, Nara (Japon), pp. 1-10. HAL : hal-01301079.
- Omar Hasan, Jingwei Miao, Sonia Ben Mokhtar & Lionel Brunie (2013). « A Privacy Preserving Prediction-based Routing Protocol for Mobile Delay Tolerant Networks ». The 27th IEEE International Conference on Advanced Information Networking and Applications (AINA-2013), 25 mars 2013, Barcelona (Espagne), pp. 546-553. doi : 10.1109/AINA.2013.6. HAL : hal-01339143.
- Pierre-Louis Aublin, Sonia Ben Mokhtar & Vivien Quema (2013). « RBFT: Redundant Byzantine Fault Tolerance ». ICDCS - International Conference on Distributed Computing Systems, 11 juillet 2013, Philadelphia (États-Unis). HAL : hal-00945798.

- Sonia Ben Mokhtar, Gauthier Berthou, Amadou Diarra, Vivien Quema & Ali Shoker (2013). « RAC: a Freerider-resilient, Scalable, Anonymous Communication Protocol ». ICDCS - International Conference on Distributed Computing Systems, 11 juillet 2013, Philadelphia (États-Unis). HAL : hal-00945795.
- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar & Lionel Brunie (2012). « A Self-Regulating Protocol for Efficient Routing in Mobile Delay Tolerant Networks ». The 5th IEEE International Conference on Digital Ecosystems and Technologies (IEEE DEST-CEE 2012), 18 juin 2012, Campione d'Italia (Italie), pp. 1-6. doi : 10.1109/DEST.2012.6227912. HAL : hal-01353129.
- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar, Lionel Brunie & Kangbin Yim (2012). « An Analysis of Strategies for Preventing Selfish Behavior in Mobile Delay Tolerant Networks ». The 6th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS 2012), 4 juillet 2012, Palermo (Italie), pp. 208-2015. doi : 10.1109/IMIS.2012.121. HAL : hal-01353130.
- Jingwei Miao, Omar Hasan, Sonia Ben Mokhtar & Lionel Brunie (2011). « An adaptive routing algorithm for mobile delay tolerant networks ». 14th International Symposium on Wireless Personal Multimedia Communications (WPMC), 3 octobre 2011, Brest (France), pp. 1-5. HAL : hal-01354555.
- Sonia Ben Mokhtar, Liam Mc Namara, Afra Mashhadi & Licia Capra (2010). « A Self-Organising Directory and Matching Service for Opportunistic Social Networking ». 3rd Workshop on Social Network Systems (SNS 2010), 13 avril 2010, Paris (France), p. 5. doi : 10.1145/1852658.1852663. HAL : hal-01381594.
- Vivien Quema, Sonia Ben Mokhtar & Alessio Pace (2010). « FireSpam: Spam Resilient Gossiping in the BAR Model ». SRDS - Symposium on Reliable Distributed Systems, 3 novembre 2010, New Delhi (Inde). HAL : hal-00945724.
- Afra Mashhadi, Sonia Ben Mokhtar & Licia Capra (2009). « Habit: Leveraging human mobility and social network for efficient content dissemination in Delay Tolerant Networks. ». IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM), 15 juin 2009, Kos Island (Grèce), p. 6. doi : 10.1109/WOWMOM.2009.5282467. HAL : hal-01437834.

Empowering Ambivalence - Supporting multiple interpretations in knowledge-based systems

Jury

Fabien GANDON
Erik MANNENS
Barry SMYTH
Debora ESTRIN
Dame WENDY HALL
Alain MILLE

Rapporteur
Rapporteur
Rapporteur
Examinatrice
Examinatrice
Examineur

Abstract

Building intelligent systems out of computers has been a continuous challenge for many computer scientists and developers. Among different paths to that goal, one that has been largely studied involves the explicit representation of knowledge, and the processing of those representations by generic reasoning engines. The advent of the Web, and then of mobile computing, has however dramatically changed the way we use computers, and with it our expectations of what such intelligent systems should be. It has also changed the means available to build them. The goal of this dissertation is to show how, in my work in the last ten years, I have been aiming at novel approaches to knowledge engineering, intending to tackle the new challenges and opportunities brought by the Web.

Knowledge-based AI has mostly developed on the premise that knowledge was rare, and as such should be made as stable as possible. A large part of our work has been trying to leverage the problems faced by any knowledge-based system when its context changes. Indeed, is not adaptability a core aspect of intelligence? But adaptive reasoning mechanisms must take into account, from the ground up, the dynamics of their knowledge base. It requires to embrace the fact that information is inherently ambivalent, that it acquires meaning (and hence becomes knowledge) only in the context of a particular problem or task. We have been pursuing a user-centered approach, where data collection and reasoning processes are as transparent as possible, and where meaning is not a pre-defined property of information, but negotiated and co-constructed with users.

I first present the theoretical framework that we have proposed to build knowledge-based systems exploiting activity traces. By capturing the inherent complexity of the user's task, this kind of knowledge allows for multiple interpretations, and hence requires a special kind of reasoning as well. Then I present a number of our works focusing on assisting a users in her task. One way is to simply present her with her traces in order to help her remembering them and sharing them with others. Another way is to use traces to detect failures and errors, and make helpful proposals for completing the task. The next chapter describes our activity related to Web technologies and Web standards. I show how the foundations of the Web accommodates and even encourages ambivalent information. As such, it allows to bridge the gap between documents, data and knowledge representations. In the next chapter, I focus on a specific class of Web documents, namely hypervideos. I present the models and tools we have proposed to process hypervideos, centered on the notion of annotation, and flexible enough to allow the emergence of new usages.

Finally, in the last chapter, to synthesize all the presented works, I propose the groundwork of a theoretical framework for knowledge representation, aimed to cope with, and account for multiple interpretations. In other words, it is

Pierre-Antoine
Champin

HDR

Soutenu le 13/06/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Pierre-Antoine.Champin@liris.cnrs.fr

an attempt to formalize ambivalent information and the dynamic reasoning processes that use them, allowing to build systems to adapt to the users, rather than forcing the users to adapt to them.

Publications

Revue internationale avec comité de lecture

- Florian Stegmaier, Werner Bailer, Tobias Bürger, Mari Carmen Suárez-Figueroa, Erik Mannens, Martin Höffernig, Pierre-Antoine Champin, Jean-Pierre Evain, Mario Döller & Harald Kosch (2013). « Unified Access to Media Metadata on the Web: Towards Interoperability Using a Core Vocabulary ». IEEE MultiMedia, vol. PP, pp. 22-29. doi : 10.1109/MMUL.2012.55. HAL : hal-01339134.
- Leila Yahiaoui, Yannick Prié, Zizette Boufaïda & Pierre-Antoine Champin (2011). « Redocumenting computer mediated activity from its traces: a model-based approach for narrative construction ». Journal of Digital Information (JoDI), vol. 12, p. 2088. HAL : hal-01354445.
- Pierre-Antoine Champin, Peter Briggs, Maurice Coyle & Barry Smyth (2010). « Coping with Noisy Search Experiences ». Knowledge-Based Systems, vol. 23, pp. 287-294. doi : 10.1016/j.knosys.2009.11.011. HAL : hal-01381440.
- Olivier Aubert, Pierre-Antoine Champin, Yannick Prié & Bertrand Richard (2008). « Canonical processes in active reading and hypervideo production ». Multimedia Systems Journal, vol. 14, pp. 427-433. doi : 10.1007/s00530-008-0132-2. HAL : hal-01514429.
- Kokou Yetongnon, Seksun Suwanmanee, Djamal Benslimane & Pierre-Antoine Champin (2006). « A Web Centric Semantic Mediation Approach for Spatial Information Systems ». Journal of Visual languages and computing (JVLC), vol. 17, pp. 1-24. HAL : hal-01536186.

Revue nationale avec comité de lecture

- Blandine Ginon, Stéphanie Jean-Daubias, Pierre-Antoine Champin & Marie Lefevre (2016). « Langage et outils pour la spécification et l'exécution d'assistance à l'utilisateur dans des applications existantes ». Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle. HAL : hal-01403052.
- Christine Michel, Marie-Thérèse Têtu-Delage, Pierre-Antoine Champin & Laetitia Pot (2016). « Stimuler la patrimonialisation socio-culturelle par des plateformes du Web ». Les Cahiers du numérique, vol. 12, n°3, pp. 31-50. doi : 10.3166/lcn.12.3.31-50. HAL : hal-01291003.
- Amélie Cordier, Marie Lefevre, Pierre-Antoine Champin, Alain Mille, Olivier Georgeon & Benoît Mathern (2014). « Connaissances et raisonnement sur les traces d'interaction ». Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle, vol. 28, pp. 375-396. doi : 10.3166/ria.28.375-396. HAL : hal-01265955.
- Pierre-Antoine Champin, Alain Mille & Yannick Prié (2013). « Vers des traces numériques comme objets informatiques de premier niveau ». Intellectica - La revue de l'Association pour la Recherche sur les sciences de la Cognition (ARCo), n°59, pp. 171-204. HAL : hal-00924203.

Conférences internationales avec comité de lecture

- Thomas Steiner, Rémi Ronfard, Pierre-Antoine Champin, Benoît Encelle & Yannick Prié (2015). « Curtains Up! Lights, Camera, Action! Documenting the Creation of Theater and Opera Productions with Linked Data and Web Technologies ». International Conference on Web Engineering ICWE 2015, 26 juin 2015, Amsterdam (Pays-Bas). HAL : hal-01159826.
- Rémi Ronfard, Benoit Encelle, Nicolas Sauret, Pierre-Antoine Champin, Thomas Steiner, Vineet Gandhi, Cyrille Migniot & Florent Thiery (2015). « Capturing and Indexing Rehearsals: The Design and Usage of a Digital Archive of Performing Arts ». Digital Heritage, 2 octobre 2015, Grenade (Espagne), pp. 533-540. doi : 10.1109/DigitalHeritage.2015.7419570. HAL : hal-01178053.
- Thomas Steiner, Hannes Mühleisen, Ruben Verborgh, Pierre-Antoine Champin, Benoît Encelle & Yannick Prié (2014). « Weaving the Web(VTT) of Data ». LDOW 2014, 8 avril 2014, Seoul (Corée du Sud), http://ceur-ws.org/Vol-1184/ldow2014_paper_11.pdf. HAL : hal-00984780.
- Blandine Ginon, Le Vinh Thai, Stéphanie Jean-Daubias, Marie Lefevre & Pierre-Antoine Champin (2014). « Adding epiphytic assistance systems in learning applications using the SEPIA system ». EC-TEL, 16 septembre 2014, Graz (Autriche), pp. 138-151. HAL : hal-01301072.
- Thomas Steiner, Pierre-Antoine Champin, Benoît Encelle & Yannick Prié (2014). « Self-Contained Semantic Hypervideos Using Web Components ». ISWC Developers Workshop 2014, 19 octobre 2014, Riva del Garda (Italie), pp. 96-101. HAL : hal-01301115.
- Blandine Ginon, Stéphanie Jean-Daubias, Pierre-Antoine Champin & Marie Lefevre (2014). « aLDEAS: a Language to Define Epiphytic Assistance Systems ». EKAW, 24 novembre 2014, Linköping (Suède), pp. 153-164. doi : 10.1007/978-3-319-13704-9_12. HAL : hal-01301088.
- Lionel Médini, Pierre-Antoine Champin, Michael Mrissa & Amélie Cordier (2014). « Towards semantic resource mashups ». Services and Applications over Linked APIs and Data (SALAD), workshop at ESWC, 26 mai 2014, Heraklion (Grèce), pp. 6-9. HAL : hal-01301051.
- Blandine Ginon, Stéphanie Jean-Daubias, Pierre-Antoine Champin & Marie Lefevre (2014). « Setup of epiphytic assistance systems with SEPIA ». EKAW, 23 novembre 2014, Linköping (Suède), pp. 1-4. HAL : hal-01301121.
- Pierre-Antoine Champin (2013). « RDF-REST: A Unifying Framework for Web APIs and Linked Data ». Services and Applications over Linked APIs and Data (SALAD), workshop at ESWC, 26 mai 2013, Montpellier (FR) (France), pp. 10-19. HAL : hal-00921662.
- Fares Belhadj, Vincent Boyer, Guilain Delmas, Myriam Lamolle, Chan Le Duc, Pierre-Antoine Champin & Christine Michel (2013). « Learning Café: a semantic multimedia collaborative platform for e-learning. ». IEEE International Workshop on Multimedia Technologies for E-Learning (MTEL) - IEEE International Symposium of multimedia (ISM 2013), 11 décembre 2013, Anaheim (États-Unis), 6p. HAL : hal-00905201.
- Alain Mille, Pierre-Antoine Champin, Amélie Cordier, Olivier Georgeon & Marie Lefevre (2013). « Trace-Based Reasoning - Modeling interaction traces for reasoning on experiences ». The 26th International FLAIRS Conference, 22 mai 2013 (États-Unis), pp. 1-15. HAL : hal-00830444.

- Amélie Cordier, Marie Lefevre, Pierre-Antoine Champin & Alain Mille (2013). « Modéliser les traces d'interaction pour raisonner à partir de l'expérience tracée ? ». IC - 24èmes Journées francophones d'Ingénierie des Connaissances, 5 juillet 2013, Lille (France). HAL : hal-01103671.
- Blandine Ginon, Pierre-Antoine Champin & Stéphanie Jean-Daubias (2013). « Collecting fine-grained use traces in any application without modifying it ». workshop EXPORT from the conference ICCBR, 8 juillet 2013, New York (États-Unis), pp. 1-10. HAL : hal-01339195.
- Blandine Ginon, Stéphanie Jean-Daubias & Pierre-Antoine Champin (2013). « Mise en place d'un système d'assistance personnalisée dans une application existante ». IC - 24èmes Journées francophones d'Ingénierie des Connaissances, 5 juillet 2013, Lille (France). HAL : hal-01107340.
- Raafat Zarka, Pierre-Antoine Champin, Amélie Cordier, Elod Egyed-Zsigmond, Luc Lamontagne & Alain Mille (2013). « TStore: A Trace-Base Management System using Finite-State Transducer Approach for Trace Transformation ». International Conference on Model-Driven Engineering and Software Development (MODELSWARD 2013), 19 février 2013, Barcelona (Espagne), pp. 117-122. HAL : hal-01351699.
- Vianney Le Clément De Saint-Marcq, Yves Deville, Christine Solnon & Pierre-Antoine Champin (2012). « Castor: a Constraint-based SPARQL Engine with Active Filter Processing ». 9th Extended Semantic Web Conference (ESWC), 27 mai 2012, Heraklion, Crète (Grèce), pp. 391-405. doi : 10.1007/978-3-642-30284-8_33. HAL : hal-01352967.
- Blandine Ginon, Pierre-Antoine Champin & Stéphanie Jean-Daubias (2012). « Taking into account users' knowledge, abilities and preferences to personalize animated assistant agents ». International Workshop on Personalization Approaches in Learning Environments (PALE) - Conference User Modeling, Adaptation, and Personalization, 16 juin 2012, Montreal (Canada), pp. 29-34. HAL : hal-01353006.
- Olivier Curé, Yannick Prié & Pierre-Antoine Champin (2012). « A Knowledge-Based Approach to Augment Applications with Interaction Traces ». EKAW 2012, 12 octobre 2012, Galway (Irlande), pp. 317-326. doi : 10.1007/978-3-642-33876-2_28. HAL : hal-00738349.
- Pierre-Antoine Champin, Amélie Cordier, Elise Lavoué, Marie Lefevre & Hala Skaf-Molli (2012). « User assistance for collaborative knowledge construction ». Workshop on Semantic Web Collaborative Spaces (SWCS), in conjunction with the World Wide Web 2012 International Conference, 17 avril 2012, Lyon (France), pp. 1065-1074. HAL : hal-00692091.
- Raafat Zarka, Pierre-Antoine Champin, Amélie Cordier, Elod Egyed-Zsigmond, Luc Lamontagne & Alain Mille (2012). « TStore: A Web-Based System for Managing, Transforming and Reusing Traces ». ICCBR 2012 TRUE and Story Cases Workshop, 3 septembre 2012, Lyon (France), pp. 173-182. HAL : hal-01353131.
- Benoît Encelle, Pierre-Antoine Champin, Yannick Prié & Olivier Aubert (2011). « Models for video enrichment ». Document Engineering 2011 (DocEng 2011), 19 septembre 2011, Mountain View, CA (États-Unis), pp. 85-88. doi : 10.1145/2034691.2034710. HAL : hal-01354516.
- José Francisco Saray Villamizar, Benoît Encelle, Yannick Prié & Pierre-Antoine Champin (2011). « An Adaptive Videos Enrichment System Based On Decision Trees For People With Sensory Disabilities ». 8th International Cross-Disciplinary Conference on Web Accessibility (W4A 2011), 28 mars 2011, Hyderabad (Inde), p. 7. doi : 10.1145/1969289.1969299. HAL : hal-01354397.

- Pierre-Antoine Champin & Alexandre Passant (2010). « SIOC in Action — Representing the Dynamics of Online Communities ». 6th International Conference on Semantic Systems (I-Semantics), 1 septembre 2010, Graz (Autriche), p. 12. doi : 10.1145/1839707.1839722. HAL : hal-01381517.
- Fabrizio Orlandi, Alexandre Passant & Pierre-Antoine Champin (2010). « Semantic Representation of Provenance in Wikipedia ». Second International Workshop on Role of Semantic Web in Provenance Management (SWPM 2010 - workshop of ISWC 2010), 7 novembre 2010, Shanghai (Chine), pp. 1-6. HAL : hal-01381598.
- Pierre-Antoine Champin, Benoît Encelle, Nicholas W. D. Evans, Magali Ollagnier-Beldame, Yannick Prié & Raphaël Troncy (2010). « Towards Collaborative Annotation for Video Accessibility ». 7th International Cross-Disciplinary Conference on Web Accessibility (W4A 2010), 26 avril 2010, Raleigh (États-Unis), p. 4. doi : 10.1145/1805986.1806010. HAL : hal-01381458.
- Pierre-Antoine Champin, Peter Briggs, Maurice Coyle & Barry Smyth (2009). « Coping with Noisy Search Experiences ». 29th SGAI International Conference on Artificial Intelligence (AI-2009), 15 décembre 2009, Cambridge (Royaume-Uni), pp. 5-18. doi : 10.1007/978-1-84882-983-1_1. HAL : hal-01437779.
- Barry Smyth & Pierre-Antoine Champin (2009). « The Experience Web: A Case-Based Reasoning Perspective ». Grand Challenges for reasoning from experiences, Workshop at IJCAI'09, 11 juillet 2009, Pasadena, CA (États-Unis), p. 6. HAL : hal-01437666.
- Pierre-Antoine Champin (2009). « T4R: Lightweight presentation for the Semantic Web ». Scripting for the Semantic Web, workshop at ESWC 2009, 31 mai 2009, Crete (Grèce), p. 12. HAL : hal-01437658.
- Barry Smyth, Pierre-Antoine Champin, Peter Briggs & Maurice Coyle (2009). « The Case-Based Experience Web ». WebCBR workshop at ICCBR'09, 20 juillet 2009, Seattle (États-Unis), inconnue. HAL : hal-01437729.
- Lotfi Sofiane Settouti, Yannick Prié, Damien Cram, Pierre-Antoine Champin & Alain Mille (2009). « A Trace-Based Framework for supporting Digital Object Memories ». 1st International Workshop on Digital Object Memories (DOMe'09) in the 5th International Conference on Intelligent Environments (IE 09), 19 juillet 2009, Barcelona (Espagne), pp. 39-44. doi : 10.3233/978-1-60750-056-8-39. HAL : hal-01437735.
- Fadi Badra, Rokia Bendaoud, Rim Bentebibel, Pierre-Antoine Champin, Julien Cojan, Amélie Cordier, Sylvie Després, Stéphanie Jean-Daubias, Jean Lieber, Thomas Meilender et al. (2008). « TAAABLE: Text Mining, Ontology Engineering, and Hierarchical Classification for Textual Case-Based Cooking ». 9th European Conference on Case-Based Reasoning - ECCBR 2008, Workshop Proceedings, 1 septembre 2008, Trier (Allemagne), pp. 219-228. HAL : inria-00337666.
- Mahmoud Barhamgi, Pierre-Antoine Champin & Djamal Benslimane (2007). « Web services-based query rewriting and resolution in loosely coupled information systems ». 3rd International Conference on Web Information Systems and Technologies (WEBIST 2007), 3 mars 2007, Barcelona (Espagne), pp. 92-96. HAL : hal-01536267.
- Mahmoud Barhamgi, Pierre-Antoine Champin, Djamal Benslimane & Aris Ouksel (2007). « Composing Data-Providing Web Services in P2P-based Collaboration Environments ». 19th International Conference on Advanced Information Systems Engineering (CAISE'07), 11 juin 2007, Trondheim (Norvège), pp. 513-545. HAL : hal-01536268.

- Mahmoud Barhamgi, Pierre-Antoine Champin & Djamel Benslimane (2007). « A Framework for Service-Based Query Resolution in eEnvironments ». 3rd IEEE International Workshop on Databases for Next-Generation Researchers (SWOD'07) at ICDE, 15 avril 2007, Istanbul (Turquie), pp. 97-102. doi : 10.1109/SWOD.2007.353205. HAL : hal-01536264.
- Pierre-Antoine Champin & Yannick Prié (2007). « Models for sustaining emergence of practices for hypervideo ». Workshop On Semantically Aware Document Processing And Indexing (SADPI 07), 22 mai 2007, Montpellier (France), pp. 1-10. HAL : hal-01501884.
- Pierre-Antoine Champin, Geert-Jan Houben & Philippe Thiran (2007). « Cross: an OWL wrapper for reasoning with relational databases ». 26th International Conference on Conceptual Modeling (ER 2007), 5 novembre 2007, Auckland (Nouvelle-Zélande), pp. 502-517. doi : 10.1007/978-3-540-75563-0_34. HAL : hal-01582692.
- Pierre-Antoine Champin (2007). « Representing data as resources in RDF and OWL ». ICDT Workshop on Emerging Research Opportunities in Web Data Management (EROW 2007), 13 janvier 2007, Barcelona (Espagne), pp. 1-9. HAL : hal-01501729.
- Olivier Aubert, Pierre-Antoine Champin & Yannick Prié (2006). « Integration of Semantic Web Technology in an Annotation-based Hypervideo System ». First International Workshop on Semantic Web Annotations for Multimedia (SWAMM 2006), 22 mai 2006, Edinburgh (Royaume-Uni). HAL : hal-01583149.
- Mahmoud Barhamgi, Djamel Benslimane & Pierre-Antoine Champin (2006). « A framework for data and Web services semantic mediation in Peer-to-Peer based Medical Information Systems ». 19th IEEE Symposium on Computer-Based Medical Systems (CBMS 2006), 22-23 June 2006. IEEE Computer Society, 22 juin 2006, Utah (États-Unis), pp. 87-92. HAL : hal-01536198.
- Seksun Suwanmanee, Djamel Benslimane, Pierre-Antoine Champin & Philippe Thiran (2005). « Wrapping And Integrating Heterogeneous Databases With OWL ». 7th International Conference on Enterprise Information Systems, (ICIES 2005), 28 mai 2005, Miami, Florida, USA (États-Unis), pp. 11-18. HAL : hal-01536164.
- Olivier Aubert, Pierre-Antoine Champin & Yannick Prié (2004). « Advène, une plateforme pour l'édition et la visualisation d'hypervidéos ». Journées Modèles documentaires de l'audiovisuel, 22 juin 2004, La Rochelle (France). HAL : sic_00001270.
- Rocio Abascal, Michel Beigbeder, Aurélien Bénel, Sylvie Calabretto, Bertrand Chabbat, Pierre-Antoine Champin, Nouredine Chatti, David Jouve, Yannick Prié, Béatrice Rumpler et al. (2003). « Modéliser la structuration multiple des documents ». H2PTM'2003, Hypermedias, Hypertexts, Products, Tools and Methods, 26 septembre 2003, Paris (France), 4p. HAL : emse-00949192.
- Rocio Abascal, Michel Beigbeder, Aurélien Bénel, Sylvie Calabretto, Bertrand Chabbat, Pierre-Antoine Champin, Nouredine Chatti, David Jouve, Yannick Prié, Béatrice Rumpler et al. (2003). « Documents à structures multiples ». SETIT'2004, Sciences of Electronic, Technology of Information and Telecommunications, - 20 mars 2003, Sousse (Tunisie). HAL : emse-00949199.
- Pierre-Antoine Champin & Christine Solnon (2003). « Measuring the similarity of labeled graphs ». 5th Int. Conf. On Case-Based Reasoning (ICCB 2003), 18 juin 2003, Trondheim (NO) (Norvège), pp. 80-95. doi : 10.1007/3-540-45006-8_9. HAL : hal-01541469.
- Pierre-Antoine Champin (2003). « ARDECO: an assistant for experience reuse in CAD ». From Structured Cases to Unstructured Problem Solving Episodes

For Experience-Based Assistance (Workshop 5 of ICCBR'03), 23 juin 2003, Trondheim (Norvège). HAL : hal-01583142.

- Pierre-Antoine Champin, Yannick Prié & Alain Mille (2003). « MUSETTE: Modeling USEs and Tasks for Tracing Experience ». Workshop 5 'From Structured Cases to Unstructured Problem Solving Episodes For Experience-Based Assistance', ICCBR'03, 23 juin 2003, Trondheim (Norvège), pp. 279-286. HAL : hal-01583134.

Conférences nationales avec comité de lecture

- Rémi Casado, Nathalie Guin, Pierre-Antoine Champin & Marie Lefevre (2017). « kTBS4LA : une plateforme d'analyse de traces fondée sur une modélisation sémantique des traces ». Méthodologies et outils pour le recueil, l'analyse et la visualisation des traces d'interaction - ORPHEE-RDV, 30 janvier 2017, Font-Romeu (France). HAL : hal-01637548.
- Le Vinh Thai, Blandine Ginon, Stéphanie Jean-Daubias, Marie Lefevre & Pierre-Antoine Champin (2015). « Modèle d'articulation entre les règles définissant un système d'assistance aLDEAS ». RJCIA, 30 juin 2015, Rennes (France). HAL : hal-01267129.
- Mohamed Besnaci, Nathalie Guin & Pierre-Antoine Champin (2015). « Acquisition de connaissances pour importer des traces existantes dans un système de gestion de bases de traces ». IC2015, 3 juillet 2015, Rennes (France). HAL : hal-01164384.
- Bryan Kong Win Chang, Marie Lefevre, Nathalie Guin & Pierre-Antoine Champin (2015). « SPARE-LNC : un langage naturel contrôlé pour l'interrogation de traces d'interactions stockées dans une base RDF ». IC2015, 3 juillet 2015, Rennes (France). HAL : hal-01164383.
- Fatma Derbel, Pierre-Antoine Champin, Amélie Cordier & Damien Munch (2015). « Authentification d'un utilisateur à partir de ses traces d'interaction ». Treizièmes Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA 2015), 30 juin 2015, Rennes (France). HAL : hal-01178926.
- Blandine Ginon, Stéphanie Jean-Daubias, Pierre-Antoine Champin & Marie Lefevre (2014). « aLDEAS : un langage de définition de systèmes d'assistance épiphytes (Prix du meilleur article d'IC2014) ». IC - 25èmes Journées francophones d'Ingénierie des Connaissances, 16 mai 2014, Clermont-Ferrand (France), pp. 137-148. HAL : hal-01015325.
- Pierre-Antoine Champin, C. Michel & Magali Ollagnier-Beldame (2014). « Des traces à la connaissance à l'ère du Web ». Table ronde « Informatique » – Quelle place tient la 'trace' dans les nouvelles formes d'apprentissage médié par un environnement numérique ? Objet utilitaire ? Objet premier ? Séminaire de l'ARCO, Institut des Sciences de la Communication, Paris (France). HAL : halshs-01518817.
- Vianney Le Clément De Saint-Marcq, Yves Deville, Christine Solnon & Pierre-Antoine Champin (2012). « Un solveur léger efficace pour interroger le Web Sémantique ». JFPC 2012, 24 mai 2012, Toulouse (France). HAL : hal-00809859.
- Blandine Ginon, Stéphanie Jean-Daubias & Pierre-Antoine Champin (2012). « Une assistance générique pour les utilisateurs d'applications fortement orientées connaissances ». 23es Journées Francophones d'Ingénierie des Connaissances - IC 2012, 29 juin 2012, Paris (France), pp. 167-174. HAL : hal-00816683.
- Pierre-Antoine Champin, Amélie Cordier, Sylvie Després, Béatrice Fuchs, Jean Lieber & Alain Mille (2008). « Construction manuelle de la partie haute d'une

ontologie modulaire destinée à une annotation de cas textuels - étude de cas pour une application culinaire dans le cadre du projet Taaable ». 16ème atelier de Raisonnement à Partir de Cas, 2 avril 2008, Nancy (France), pp. 36-49. HAL : hal-01501252.

- Olivier Aubert, Pierre-Antoine Champin & Yannick Prié (2008). « Advène, une plate-forme ouverte pour la construction d’hypervidéos ». Ludovia 2008, 29 août 2008, Ax-les-Thermes (France). HAL : hal-01515529.
- Julien Laflaquière, Pierre-Antoine Champin, Yannick Prié & Alain Mille (2005). « Approche de modélisation de l’expérience : utilisation de systèmes complexes pour l’assistance aux tâches de veille informatiquement médiées ». ISKO-France 2005, 28 avril 2005, INIST-CNRS, Vandoeuvre-lès-Nancy (France), pp. 209-230. HAL : hal-01583146.
- Pierre-Antoine Champin, Yannick Prié & Alain Mille (2004). « MUSETTE : a framework for Knowledge from Experience ». Extraction et gestion des connaissances (EGC’2004) (article court) (France), pp. 129-134. HAL : hal-00824598.
- Sébastien Sorlin, Pierre-Antoine Champin & Christine Solnon (2003). « Mesurer la similarité de graphes étiquetés ». 9èmes Journées Nationales sur la résolution pratique de problèmes NP-Complets (JNPC 2003) - pages 325-339 Juin 2003, 1 juin 2003, Amiens (France), pp. 91-107. HAL : hal-01541503.
- Olivier Aubert, Pierre-Antoine Champin & Yannick Prié (2003). « Instrumentation de documents audiovisuels : temporalisations et détemporalisation dans le projet Advène ». Workshop “ Temps et documents numériques”, 19 novembre 2003, Grenoble (France). HAL : hal-01583140.

Autres

- Blandine Ginon, Stéphanie Jean-Daubias & Pierre-Antoine Champin (2013). « Adjonction de systèmes d’assistance personnalisée à des EIAH existants ». EIAH 2013 - Environnements Informatiques pour l’Apprentissage Humain, 28 mai 2013, Toulouse (France). HAL : hal-01339171.
- Vianney Le Clément De Saint-Marcq, Yves Deville, Christine Solnon & Pierre-Antoine Champin (2011). « Castor: Using Constraint Programming to Solve SPARQL Queries ». 10th International Semantic Web Conference (ISWC), Poster session, 23 octobre 2011, Bonn (Allemagne). HAL : hal-01354538.

Éditions scientifique d’ouvrages

- Pierre-Antoine Champin (2017). « Le Web de données ». HAL : cel-01561241.

Contributions à un ouvrage

- Musette: uses-based annotation for the Semantic Web. P-A. Champin, Y. Prié. Annotation for the Semantic Web 2003.

Jury

Loïc BARTHE, PR, Université Paul Sabatier, Toulouse
Atilla BASKURT, PR, INSA Lyon
Raphaëlle CHAINE, PR, Université Claude Bernard Lyon 1
Jean-Michel DISCHLER, PR, Université de Strasbourg
Éric GALIN, PR, Université Claude Bernard Lyon 1

Résumé

Les mondes virtuels sont de plus en plus présents dans de nombreuses applications telles que les effets spéciaux, les jeux vidéo ou les simulateurs. Les terrains jouent un rôle principal dans ces mondes virtuels car ils en sont le socle. De nombreux travaux de recherche se sont penchés sur la modélisation, la génération ou l'édition de ces terrains.

Dans cet exposé, j'identifierai les défis majeurs associés à ces travaux et ferai ensuite une analyse des tendances de l'état de l'art face à ces défis.

Je présenterai ensuite plusieurs de mes contributions dans le domaine de la modélisation de terrains virtuels. Ces contributions font apparaître une amélioration dans le domaine de l'étendue et la précision des terrains générés, ainsi que dans le contrôle qu'offre le modèle à l'utilisateur. Je commencerai par une contribution dans le domaine des fractales, et la manière dont on peut insérer des informations de détails pour rompre le caractère systématique et aléatoire des fractales. Je montrerai ensuite comment il est possible de détourner la classique équation de diffusion de la chaleur pour en faire un modèle de terrains dont les caractéristiques fines sont facilement contrôlables. J'introduirai ensuite un modèle de représentation de terrains par un arbre de construction qui permet de combiner des portions de terrains de natures variées en un seul modèle. Enfin, je présenterai un modèle de terrains qui repose sur une représentation parcimonieuse, combinant des éléments d'un dictionnaire d'atomes de terrains ainsi que les nombreuses applications que ce modèle offre en synthèse, problème inverse ou en amplification de terrains.

Je clôturerai par un exposé des pistes de recherches qui restent à explorer et que je juge prometteuses telles que le deep learning ainsi que la stratégie associée.

Publications

Revue internationale avec comité de lecture

- Antoine Webanck, Yann Cortial, Eric Guérin & Eric Galin (2018). « Procedural Cloudscapes ». Computer Graphics Forum. HAL : hal-01730789.
- Guillaume Cordonnier, Eric Galin, James Gain, Bedrich Benes, Eric Guérin, Adrien Peytavie & Marie-Paule Cani (2017). « Authoring Landscapes by Combining Ecosystem and Terrain Erosion Simulation ». ACM Transactions on Graphics. doi : 10.1145/3072959.3073667. HAL : hal-01518967.
- Oscar Argudo, Carlos Andujar, Antonio Chica, Eric Guérin, Julie Digne, Adrien Peytavie & Eric Galin (2017). « Coherent multi-layer landscape synthesis ». Visual Computer. doi : 10.1007/s00371-017-1393-6. HAL : hal-01520366.
- Eric Guérin, Julie Digne, Eric Galin, Adrien Peytavie, Christian Wolf, Bedrich Benes & Benoît Martinez (2017). « Interactive Example-Based Terrain

Eric Guérin

HDR

Soutenu le 14/11/2017

Établissement :
INSA Lyon

Contact :
Eric.Guerin@liris.cnrs.fr

Authoring with Conditional Generative Adversarial Networks ». *Transactions on Graphics (Proceedings of Siggraph Asia 2017)*, vol. 36, pp. 228-228. doi : 10.1145/3130800.3130804. HAL : hal-01583706.

- Eric Guérin, Eric Galin, François Grosbellet, Adrien Peytavie & Jean-David Genevaux (2016). « Efficient modeling of entangled details for natural scenes ». *Computer Graphics Forum*, vol. 35, n°7, pp. 257-267. doi : 10.1111/cgf.13023. HAL : hal-01370684.
- Eric Guérin, Julie Digne, Eric Galin & Adrien Peytavie (2016). « Sparse representation of terrains for procedural modeling ». *Computer Graphics Forum*, vol. 35, n°2, pp. 177-187. doi : 10.1111/cgf.12821. HAL : hal-01258986.
- Guillaume Cordonnier, Jean Braun, Marie-Paule Cani, Bedrich Benes, Eric Galin, Adrien Peytavie & Eric Guérin (2016). « Large Scale Terrain Generation from Tectonic Uplift and Fluvial Erosion ». *Computer Graphics Forum*, vol. 35, n°2, pp. 165-175. doi : 10.1111/cgf.12820. HAL : hal-01262376.
- François Grosbellet, Adrien Peytavie, Eric Guérin, Eric Galin, Stéphane Mérillou & Bedrich Benes (2015). « Environmental Objects for Authoring Procedural Scenes ». *Computer Graphics Forum*, vol. 35, n°1, pp. 296-308. doi : 10.1111/cgf.12726. HAL : hal-01250526.
- Jean-David Genevaux, Eric Galin, Adrien Peytavie, Eric Guérin, Cyril Briquet, François Grosbellet & Bedrich Benes (2015). « Terrain Modelling from Feature Primitives ». *Computer Graphics Forum*, vol. 34, n°6, pp. 198-210. doi : 10.1111/cgf.12530. HAL : hal-01257198.
- Jean-David Genevaux, Eric Galin, Eric Guérin, Adrien Peytavie & Bedrich Benes (2013). « Terrain Generation Using Procedural Models Based on Hydrology ». *ACM Transactions on Graphics*, vol. 32, 143:1-143:13. doi : 10.1145/2461912.2461996. HAL : hal-01339224.
- Eric Galin, Adrien Peytavie, Eric Guérin & Bedrich Benes (2011). « Authoring Hierarchical Road Networks ». *Computer Graphics Forum*, vol. 30, pp. 2021-2030. doi : 10.1111/j.1467-8659.2011.02055.x. HAL : hal-01354487.
- Nicolas Maréchal, Eric Guérin, Eric Galin, Stéphane Mérillou & Nicolas Mérillou (2010). « Heat Transfer Simulation for Modeling Realistic Winter Sceneries ». *Computer Graphics Forum*, vol. 29, n°2, 10 p. HAL : hal-00475643.
- Houssam Hnaidi, Eric Guérin, Samir Akkouche, Adrien Peytavie & Eric Galin (2010). « Feature based terrain generation using diffusion equation ». *Computer Graphics Forum*, vol. 29, pp. 2179-2186. doi : 10.1111/j.1467-8659.2010.01806.x. HAL : hal-01381590.
- Houssam Hnaidi, Eric Guérin & Samir Akkouche (2010). « Multiresolution control of curves and surfaces with a self-similar model ». *Fractals*, vol. 18, pp. 271-286. HAL : hal-01381459.
- Eric Galin, Adrien Peytavie, Eric Guérin & Nicolas Marechal (2010). « Procedural Generation of Roads ». *Computer Graphics Forum*, vol. 29, pp. 429-438. HAL : hal-01381447.

Conférences internationales avec comité de lecture

- Leandro Cruz, Luiz Velho, Eric Galin, Adrien Peytavie & Eric Guérin (2015). « Patch-based Terrain Synthesis ». *International Conference on Computer Graphics Theory and Applications*, 14 mars 2015, Berlin (France). doi : 10.5220/0005360201890194. HAL : hal-01713220.
- Leandro Cruz, Luiz Velho, Djalma Lucio, Eric Galin, Adrien Peytavie & Eric Guérin (2014). « Landscape Specification Resizing ». *CLEI*, 15 septembre 2014, Montevideo (Uruguay), pp. 1-10. HAL : hal-01301096.

- Nicolas Marechal, Eric Galin, Eric Guérin & Samir Akkouche (2010). « Component-Based Model Synthesis for Low Polygonal Models ». *Graphics Interface*, 31 mai 2010, Ontario (Canada), pp. 217-224. HAL : hal-01381461.
- Houssam Hnaidi, Eric Guérin & Samir Akkouche (2008). « Fractal/Wavelet representation of objects ». *The International Conference on Information & Communication Technologies: from Theory to Applications - ICTTA'08*, 11 avril 2008, Damascus (Syrie), pp. 1-5. HAL : hal-01500841.
- Eric Guérin, Eric Tosan & Atilla Baskurt (2003). « Fractal Compression of Images with Projected IFS ». *Int. Picture Coding Symposium, PCS'03*, Saint-Malo, France, 2003., 23 avril 2003, Saint Malo (France). HAL : hal-01581315.
- Eric Guérin, Eric Tosan & Atilla Baskurt (2003). « Fractal approximation and compression using projected ifs ». *IATA'2003, Interdisciplinary Approaches in Fractal Analysis*, Bucharest, Romania, 1 janvier 2003, Bucharest (Roumanie). HAL : hal-01581320.

Conférences nationales avec comité de lecture

- Jean-David Génevaux, François Grosbellet, Eric Galin, Adrien Peytavie, Eric Guérin, Cyril Briquet & Bedřich Beneš (2014). « Modélisation de terrains par primitives ». *Journées de l'Association Française d'Informatique Graphique*, 28 novembre 2014, Reims (France). HAL : hal-01259310.
- François Grosbellet, Adrien Peytavie, Eric Guérin, Eric Galin, Stéphane Mérillou & Bedrich Benes (2013). « Changement d'apparence au cours des saisons ». *26èmes Journées de l'Association Française d'Informatique Graphique, du chapitre français d'Eurographics et du Groupement de Recherche IG (AFIG)*, 26 novembre 2013, Limoges (France), pp. 1-10. HAL : hal-01339315.
- François Grosbellet, Adrien Peytavie, Eric Guérin, Eric Galin, Stéphane Mérillou & Bedrich Benes (2013). « Changements d'apparences au cours des saisons ». *AFIG*, 26 novembre 2013, Limoges (France). HAL : hal-01250529.
- Jean-David Genevaux, Eric Galin, Eric Guérin, Adrien Peytavie & Bedrich Benes (2012). « Génération procédurale de rivières et de terrains ». *25èmes Journées de l'Association Française d'Informatique Graphique (AFIG)*, 20 novembre 2012, Calais (France), pp. 1-10. HAL : hal-01353170.
- Adrien Peytavie, Nicolas Marechal, Eric Guérin & Eric Galin (2009). « Génération procédurale de routes ». *AFIG'09 : 22e journées de l'Association Française d'Informatique Graphique*, 25 novembre 2009, Arles (France), pp. 123-132. HAL : hal-01437829.
- Nicolas Marechal, Eric Galin, Eric Guérin & Samir Akkouche (2008). « Génération de Variétés d'Objets par Fragments ». *21èmes journées de l'Association Française d'Informatique Graphique (AFIG 2008)*, 21 novembre 2008, Toulouse (France), pp. 1-9. HAL : hal-01501350.
- Houssam Hnaidi, Eric Guérin & Samir Akkouche (2007). « Insertion de détail dans des figures autosimilaires ». *20ièmes Journées de l'AFIG 2007*, Marne la Vallée (France). HAL : hal-01537641.
- Eric Tosan, Eric Guérin & Ikbal Zammouri (2005). « Modèle itératif et surfaces de subdivision ». *GTMG 2005, Journées du Groupe de Travail en Modélisation Géométrique*, 17 mars 2005, Poitiers (France). HAL : hal-01537647.
- Gilles Gouaty, Eric Tosan & Eric Guérin (2004). « Visualisation rapide de formes fractales ». *Journées de l'AFIG 2004*, Poitiers (France). HAL : hal-01537622.

Autres conférences

- François Grosbellet, Stéphane Mérillou, Eric Galin, Eric Guérin, Adrien Peytavie & Bedrich Benes (2013). « Modélisation procédurale de changements de saison ». Workshop Doctorant XLIM, 26 septembre 2013, Limoges (France). HAL : hal-01339316.

Contributions à un ouvrage

- Eric Guérin & Eric Tosan (2005). « Fractal Inverse Problem: Approximation Formulation and Differential Methods ». Fractal in Engineering: new trends in theory and applications, Jacques Lévy-Véhel, Evelyne Lutton, Springer, pp. 271-285. doi : 10.1007/1-84628-048-6_17. HAL : hal-01537632.

Edition scientifique d'ouvrages

- Jan Bender, Arjan Kuijper, Dieter Fellner & Eric Guérin (2012). « VRIPHYS 2012: Ninth Workshop on Virtual Reality Interactions and Physical Simulations ». Eurographics Association. HAL : hal-01353153.

Méthodologie de développement de logiciels dans un environnement informatiquement vert

Jury

Mourad CHABANE OUSSALAH, PR, Université de Nantes	Rapporteur
Jean-Marc PIERSON, PR, Université de Toulouse	Rapporteur
Ladjel BELLATRECHE, PR, ENSMA Poitiers	Examineur
Genoveva VARGAS-SOLAR, Chercheur, CNRS, LIG Grenoble	Examinatrice
Parisa GHODOUS, PR, Université Lyon 1	Directrice de thèse
Jean-Patrick GELAS, MC, Université Lyon 1	Co-directeur de thèse
Gulfem ISIKLAR ALPTEKIN, Université de Galatasaray	Co-directrice de thèse
Laurent LEFÈVRE, CR, INRIA, ENS Lyon	Invité

Résumé

Le nombre de périphériques mobiles (smartphone, tablette, ordinateur portable, etc.) et les internautes augmentent continuellement. En raison de l'accessibilité du cloud computing, de l'Internet et de l'Internet des Objets (IdO), les utilisateurs utilisent de plus en plus d'applications logicielles qui provoquent un effet croissant sur les émissions de gaz à effet de serre. Ainsi, les TIC (Technologies de l'Information et de la Communication) sont responsables d'environ 2% des émissions mondiales de gaz à effet de serre qui sont équivalentes à celles émises par l'industrie aérienne. Selon des rapports récents, le Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC), les émissions de CO₂ dus aux TIC augmentent rapidement. Néanmoins, les TIC, en permettant de résoudre des problèmes complexes dans d'autres secteurs, peuvent grandement et facilement participer pour réduire une partie importante des 98% restants des émissions mondiales de CO₂.

L'utilisation du logiciel implique des opérations matérielles qui sont physiquement responsables de la consommation d'énergie. Par conséquent, le logiciel est indirectement impliqué dans la consommation d'énergie. Ainsi, nous devons réduire la consommation d'énergie du logiciel tout en conservant les mêmes fonctionnalités pour le logiciel afin de créer des logiciels durables et verts.

Premièrement, dans ce travail de thèse, nous définissons les termes « durable et vert » dans le domaine du logiciel afin de créer des logiciels respectant les critères de ces termes. Pour créer un produit logiciel, nous devons suivre un processus d'ingénierie logicielle. Par conséquent, nous décrivons des critères durables et verts à respecter après chaque étape de ce processus afin d'établir un processus d'ingénierie logicielle durable et écologique.

En particulier, nous nous concentrons sur l'estimation de la consommation d'énergie du logiciel. De nombreux travaux ont essayé de proposer divers outils pour estimer la consommation d'énergie due aux logiciels afin de réduire l'empreinte carbone. Pendant longtemps, les solutions proposées se sont concentrées uniquement sur la conception du matériel, mais ces dernières années, les aspects logiciels sont également devenus importants. Malheureusement, ces études, dans la plupart des cas, ne considèrent que le CPU et négligent tous les autres composants. Les modèles de consommation d'énergie existants doivent être améliorés en tenant compte de plus de composants susceptibles de consommer de l'énergie pendant l'exécution d'une application. L'écriture d'un logiciel durable, performant et vert nécessite de comprendre le comportement de consommation d'énergie d'un programme informatique. L'un des avantages est que les développeurs, en améliorant leurs implémentations du code source, optimiseront la

Hayri Acar

Thèse

Soutenue le 23/11/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Parisa.Ghodous@liris.cnrs.fr

consommation d'énergie du logiciel. De plus, il existe un manque d'outil d'analyse pour surveiller dynamiquement la consommation d'énergie du code source de plusieurs composants. Ainsi, nous proposons GMTEEC (Modèle Générique d'Outil pour Estimer la Consommation Energétique) qui se compose de quatre couches aidant et guidant la construction d'un outil permettant d'estimer la consommation énergétique d'un logiciel. Ainsi, dans notre travail, en respectant les couches de GMTEEC, nous créons TEEC (Outil pour Estimer la Consommation Energétique) qui repose sur une formule mathématique établie pour chaque composant (CPU, mémoire, disque dur, réseau) susceptible de consommer de l'énergie afin d'estimer la consommation totale d'énergie du logiciel composée de la somme de chaque consommation d'énergie par composant. De plus, nous ajoutons à TEEC la capacité de localiser dynamiquement les points chauds qui sont les parties du code source consommant la plus grande quantité d'énergie afin d'aider et guider les développeurs à optimiser leur code source et à créer des logiciels efficaces, durables et verts.

Nous avons réalisé une variété d'expériences pour valider la précision et la qualité du processus d'ingénierie logicielle durable et verte et la précision et la qualité de TEEC. Les résultats démontrent la possibilité de réduire la consommation d'énergie et d'améliorer les performances à des coûts limités avec un important impact positif sur l'environnement.

Abstract

The number of mobile devices (smartphone, tablet, laptop, etc.) and Internet users are continually increasing.

Due to the accessibility provided by cloud computing, Internet and Internet of Things (IoT), users use more and more software applications which cause an increasing effect on gas emission. Thus, ICT (Information and Communication Technologies) is responsible of around 2% worldwide greenhouse gas emissions which is equivalent of that emitted by the airline industry. According to recent reports, the Intergovernmental Panel on Climate Change (IPCC), CO₂ emissions due to ICT are increasing widely. Nevertheless, ICT, in allowing to solve complex problems in other sectors, can greatly and easily participate to reduce significant portion of the remaining 98% of global CO₂ emissions. The use of software implies hardware operations which are physically responsible of energy consumption. Consequently, software is indirectly involved in the energy consumption. Thus, we need to reduce software energy consumption while maintaining the same functionalities for the software in order to build sustainable and green software.

Firstly, in this thesis work, we define the terms sustainable and green in the area of software development. To build a software product, we need to follow a software engineering process. Hence, we define and describe sustainable and green criteria to be respected after each step of this process in order to establish a sustainable and green software engineering process.

Then, we focus on the software energy consumption estimation. Many research works tried to propose various tools to estimate the energy consumption due to software in order to reduce carbon footprint. Unfortunately, these studies, in the majority of cases, consider only the CPU and neglects all others components. Existing power consumption models need to be improved by taking into account more components susceptible to consume energy during runtime of an application. Writing sustainable, power efficient and green software necessitates to understand the power

consumption behavior of a computer program. One of the benefits is the fact that developers, by improving their source code implementations, will optimize software power consumption. Moreover, there is a lack of analyzing tool to dynamically monitor source code energy consumption of several components. Thus, we propose GMTEEC (Generic Model of a Tool to Estimate Energy Consumption) which is composed of four layers assisting developers to build a tool estimating the software power consumption. Hence, in our work, respecting the layers of GMTEEC, we develop TEEC (Tool to Estimate Energy Consumption) which is based on mathematical formula established for each component (CPU, memory, hard disk, network) in order to estimate the total software energy consumption. Moreover, we add in TEEC the capacity to locate dynamically the hotpoints which are the parts of source code consuming the greater amount of energy in order to help and guide developers to optimize their source code and build efficient, sustainable and green software.

We performed a variety of experiments to validate the accuracy and quality of the sustainable and green software engineering process and TEEC. The results demonstrate the possibility to save significant quantity of energy and time at limited costs with an important positive impact on environment.

Publications

Revue internationale avec comité de lecture

- Hayri Acar, Gülfem Alptekin, Jean-Patrick Gelas & Parisa Ghodous (2016). « The Impact of Source Code in Software on Power Consumption ». International Journal of Electronic Business Management, vol. 14, pp. 42-52. HAL : hal-01496266.

Conférences internationales avec comité de lecture

- Hayri Acar, Gülfem Isiklar Alptekin, Jean-Patrick Gelas & Parisa Ghodous (2015). « Towards a Green and Sustainable Software ». Concurrent Engineering 2015, 22 juillet 2015, Delf (Pays-Bas), pp. 471-480. HAL : hal-01192692.

Autres

- Hayri Acar, Gülfem Isiklar Alptekin, Jean-Patrick Gelas & Parisa Ghodous (2015). « A Green approach to save energy consumed by software ». ICT4S, 9 septembre 2015, Copenhague (Danemark). HAL : hal-01197452.



Une approche générique pour la construction collaborative d'éditions numériques critiques

Jury

Elisabeth MURISASCO, PR, Université de Toulon	Rapporteuse
Ethan MUNSON, PR, Université de Wisconsin-Milwaukee	Rapporteur
Elena PIERAZZO, PR, Université Grenoble Alpes	Examinatrice
Jean-Yves VION-DURY, Docteur, Naver Labs Europe	Examineur
Sylvie CALABRETTO, PR, INSA Lyon	Directrice de thèse
Olivier FERRET, PR, Université Lumière Lyon 2	Co-directeur de thèse
Pierre-Edouard PORTIER, MC, INSA Lyon	Co-directeur de thèse

Résumé

Dans la continuité des éditions critiques traditionnelles, les années 2000 ont vu l'émergence de nouveaux objets éditoriaux : les éditions critiques numériques. De telles éditions se distinguent des numérisations de fonds patrimoniaux en offrant une transcription enrichie de données secondaires (notes et commentaires, indication des sources, données bio/bibliographiques, etc.). De tels objets prennent la forme de documents structurés, e.g. de documents XML. En pratique, les éditions critiques s'appuient sur une structure de données définie explicitement sous la forme d'un schéma. Reflet de la politique éditoriale, le schéma indique le vocabulaire et la grammaire d'annotation mis en œuvre dans l'édition.

Si XML est le standard de facto pour l'annotation textuelle, il est établi qu'une structure hiérarchique représente mal un texte littéraire, et a fortiori un texte annoté. D'autres modèles d'annotation, reposant sur des formalismes de graphes plus généraux, ont été proposés : les modèles de documents multist structurés. Cependant, les mécanismes de validation pour documents multist structurés échouent à embrasser la catégorie de graphes la plus prometteuse pour la réalisation d'annotations, à savoir des graphes cycliques, du moins avec une complexité algorithmique raisonnable.

Sur un autre plan, il apparaît que les projets d'édition mettent en œuvre un mode de travail linéaire, comme suit : un schéma est initialement proposé ; il est dans un second temps instancié, puis les documents résultant de cette instanciation sont publiés. Le problème pratique est que les éditeurs, dans la phase de conception du schéma, anticipent rarement sur l'ensemble des situations qui seront rencontrées dans les textes annotés. En d'autres termes, les schémas sont appelés à évoluer pendant la construction de l'édition. Or une mise à jour du schéma doit s'accompagner d'une mise à jour des données structurées...

Nos travaux s'articulent autour de trois axes : la représentation d'annotations critiques concurrentes ou multist structurées ; la validation de telles données ; l'assistance à la mise à jour de ces données en cas d'amendement du schéma correspondant.

1. Nous avons défini un modèle de balisage déporté (stand-off markup), nommé eAG, dans lequel l'annotation est représentée sous la forme d'un graphe cyclique. Nous avons défini et formalisé la notion de chronologie composite, permettant de faire référence aux contenus de documents comportant, typiquement, du texte et de l'image, et à rendre ce type de documents compatible avec le balisage déporté. Nous avons proposé une syntaxe de balisage pour l'annotation textuelle, nommée LeAG.

2. La validation de données dont la structure, comme dans les eAG, est un

Vincent Barreton

Thèse

Soutenue le 27/11/2017

Établissement :
INSA Lyon

Contact :
Sylvie.Calabretto@liris.cnrs.fr

graphe cyclique, est coûteuse si elle repose sur une grammaire (RelaxNG, Creole, Schex, etc.). A ce titre, nous proposons la notion de simulation en tant que mécanisme de validation. Un schéma SeAG est un graphe qui valide un eAG si il simule ce dernier. Nous avons défini un mode de représentation couplé pour les schémas et les graphes d'annotation tel que, étant donné la représentation d'un schéma, seuls des graphes d'annotation simulés par le schéma puissent être représentés, rendant possible une validation « par construction » des eAG.

3. Le troisième axe de recherche s'apparente à la problématique des transformations bidirectionnelles ou du Data Exchange dans laquelle, en fonction de correspondances entre deux schémas, les données instanciant le premier schéma sont traduites dans une forme compatible avec le second. Notre approche se base sur un petit nombre de primitives, opérations élémentaires qui s'appliquent aux schémas : suppression, insertion, substitution d'une sous-partie. Un schéma est amendé par application successive de ces primitives. Chacune de ces primitives, en outre, peut être interprétée au niveau des instances et donner une nouvelle forme d'instance, compatible avec le schéma modifié. Enfin, les transformations entre instances sont bidirectionnelles, assurant une synchronisation entre les instances de deux schémas.

Abstract

Digital Scholarly Editions are critically annotated patrimonial literary resources, in a digital form. Such editions roughly take the shape of a transcription of the original resources, augmented with critical information, that is, of structured data. In a collaborative setting, the structure of the data is explicitly defined in a schema, an interpretable document that governs the way editors annotate the original resources and guarantees they follow a common editorial policy.

Digital editorial projects classically face two technical problems. The first has to do with the expressiveness of the annotation languages, that prevents from expressing some kinds of information. The second relies in the fact that, historically, schemas of long-running digital edition projects have to evolve during the lifespan of the project. However, amending a schema implies to update the structured data that has been produced, which is done either by hand, by means of ad-hoc scripts, or abandoned by lack of technical skills or human resources.

In this work, we define the theoretical ground for an annotation system dedicated to scholarly edition. We define eAG, a stand-off annotation model based on a cyclic graph model, enabling the widest range of annotation. We define a novel schema language, SeAG, that permits to validate eAG documents on-the-fly, while they are being manufactured. We also define an inline markup syntax for eAG, reminiscent of the classic annotation languages like XML, but retaining the expressivity of eAG. Eventually, we propose a bidirectional algebra for eAG documents so that, when a SeAG S is amended, giving S' , an eAG I validated by S is semi-automatically translated into an eAG I' validated by S' , and so that any modification applied to I (resp. I') is semi-automatically propagated to I' (resp. I) – hence working as an assistance tool for the evolution of SeAG schemas and eAG annotations.

Publications

Revue internationale avec comité de lecture

- Alexandre Guilbaud, Irène Passeron, Vincent Barrellon & Olivier Ferret (2014). « Editer l'Encyclopédie au 21e siècle : un projet d'édition numérique, critique et collaborative ». *Dix-Huitième Siècle*, n°46, pp. 153-166. HAL : hal-01258324.
- Alexandre Guilbaud, Irène Passeron, Marie Leca-Tsiomis, Olivier Ferret & Vincent Barrellon (2013). « Entrer dans la forteresse” : pour une édition numérique collaborative et critique de l'Encyclopédie (projet ENCCRE) ». *Recherches sur Diderot et sur l'Encyclopédie*, n°48, pp. 225-261. HAL : hal-00923750.
- Alexandre Guilbaud, Irène Passeron, Marie Leca-Tsiomis, Olivier Ferret, Vincent Barrellon & Sumi Yoichi (2013). « Entrer dans la forteresse” : pour une édition numérique collaborative et critique de l'Encyclopédie (projet ENCCRE) ». *Recherches sur Diderot et sur l'Encyclopédie*, vol. RDE, n°48, pp. 225-261. HAL : hal-01258346.

Conférences internationales avec comité de lecture

- Vincent Barrellon, Olivier Ferret, Sylvie Calabretto & Pierre-Edouard Portier (2017). « Linear Extended Annotation Graphs ». *2017 ACM Symposium on Document Engineering*, 4 septembre 2017, La Valette (Malte). HAL : hal-01636884.
- Vincent Barrellon, Pierre-Edouard Portier, Sylvie Calabretto & Olivier Ferret (2016). « Schema-aware Extended Annotation Graphs ». *ACM Symposium on Document Engineering 2016*, 16 septembre 2016, Vienne (Autriche), pp. 45-54. doi : 10.1145/2960811.2960816. HAL : hal-01481320.



Decomposition and Domination of Some Graphs

Jury

Grigori DANIELA, PR, Université de Paris Dauphine	Examinatrice
Baril JEAN-LUC, PR, Université de Bourgogne	Rapporteur
Nourine LHOUARI, PR, Université Blaise Pascal Clermont-Ferrand	Rapporteur
Zagaglia Salvi NORMA, PR, Ecole polytechnique de Milan	Examinatrice
Benbernou SALIMA, PR, Université de Paris Descartes	Examinatrice
Kheddouci HAMAMACHE, PR, Université Lyon 1	Directeur de thèse
Haddad MOHAMMED, MC, Université Lyon 1	Co-directeur de thèse

Résumé

La théorie des graphes est considérée comme un vaste champ qui permet d'explorer différentes techniques de preuve de la mathématique discrète.

Dans cette optique, nous proposons, dans cette thèse, de mettre l'accent sur trois problèmes de graphes, à savoir la décomposition de multigraphes, la $[1,2]$ -domination et le monitoring des arêtes. Ainsi, le fait d'explorer, dans ce travail de thèse, trois problèmes de graphes plus en moins distincts dans des classes de graphes différentes, nous a permis de développer plusieurs techniques de preuve ainsi qu'une multitude de façon à aborder un problème.

La première partie de cette thèse touche un aspect très important de la théorie des graphes, appelé la décomposition des graphes. Intuitivement, une décomposition en sous-graphe permet de représenter le graphe d'origine par un ensemble de copies du sous-graphe, où chaque arête du graphe initial appartient à une et une seule copie du sous-graphe. Dans cette partie, on s'intéresse plus particulièrement à la décomposition multiple d'un multigraphe complet en étoiles et cycles de même taille. Dans ce sens, des preuves formelles sont présentées pour déterminer les conditions nécessaires et suffisantes que doit avoir le multigraphe complet pour qu'une telle décomposition existe. Les deux autres parties de cette thèse, les parties les plus consistantes, abordent un problème suscitant le plus d'attention actuellement, qui est l'étude de la domination dans les graphes. Le problème original de domination consiste à trouver un ensemble de sommets (de taille minimum) dominant le reste des sommets d'un graphe. De nombreuses variantes d'intérêt à la fois théoriques et pratiques ont été proposées et étudiées dans la littérature. Dans cette partie de thèse et celle qui suit, nous nous sommes intéressés à deux variantes de domination. La première variante, appelée $[i,j]$ -domination dans les graphes, a été introduite par Chellali et al. En 2013. En plus de ses propriétés de domination, la particularité de cette variante est que chaque sommet non dominant doit être adjacent à un ensemble dominant, de taille comprise entre i et j . Plus particulièrement, nous nous sommes intéressés à la $[1,2]$ -domination. Il convient de souligner qu'il a été démontré que le problème reste NP-complet. Dans ce sens, nous avons étudié ce paramètre dans des graphes particuliers, tels que les graphes de Petersen Généralisé, ce qui rend ce problème tout aussi intéressant. Introduite par Watkins, cette famille de graphes possède un nombre de propriétés très intéressantes. D'ailleurs, plusieurs paramètres de graphes ont été étudiés sur cette classe de graphes de par sa structure qui est assez particulière. De plus, une étude de la $[1,2]$ -total domination sur cette classe de graphes est aussi menée dans cette thèse. La deuxième et dernière variante étudiée est, aussi une variante de la domination, appelée monitoring des arêtes, a été introduite par Dong et al. en 2008. Elle consiste à trouver un ensemble de sommets qui monitore (domine) l'ensemble des arêtes dans un graphe sachant qu'un sommet monitore une arête s'il forme un triangle

Fairouz Beggas

Thèse

Soutenue le 28/03/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Hamamache.Kheddouci@iris.cnrs.fr

avec les deux extrémités de l'arête. Une arête peut être monitorée par un ou plusieurs sommets. Dans ce contexte, plusieurs variantes du monitoring des arêtes sont considérées dans cette partie à savoir monitoring des arêtes, monitoring uniforme des arêtes et monitoring pondéré des arêtes. L'essence de ce problème réside dans sa nature combinatoire ainsi que son domaine d'applications, plus particulièrement dans les réseaux de capteurs sans fil. De plus, il a été prouvé que trouver un ensemble minimum pour ce problème est NP-difficile. Vu la complexité de ce type de problème, nous nous sommes intéressés, en premier temps, par une étude théorique du problème: différentes variantes, les bornes, caractérisations, etc. Par la suite, nous avons étudié le problème en profondeur dans différentes classes de graphes.

Abstract

Graph theory is considered as a field exploring a large variety of proof techniques in discrete mathematics. In this thesis, three major problems are considered: the multidecomposition of multigraphs, the $[1, 2]$ -domination and the edge monitoring. The fact that these three problems are of different nature allowed us to explore several proof techniques in this thesis.

The first part of this thesis deals with a popular aspect of research in graph theory called graph decomposition. Intuitively, a decomposition into subgraphs allows us to describe the original graph with a set of copies of these subgraphs. In this part, we investigate the problem of multidecomposition of the complete multigraph and give necessary and sufficient conditions for such a multidecomposition to exist. The second and third parts are the most important parts in terms of effort and spent time. They are devoted to problems related to domination in graphs. The original domination problem is to find a minimum set of vertices such that every vertex outside the dominating set is adjacent to at least one vertex from the dominating set. Many variants of theoretical and practical interest have been studied in the literature. The second studied problem is called the $[i, j]$ -domination in graphs. This problem was introduced by Chellali et al. in 2013. In addition to the properties of domination, this variant has the particularity that each non-dominating vertex should be adjacent to at least i dominating vertices but also to at most j of them. We particularly focus on the $[1, 2]$ -domination. It has been shown that the problem remains NP-complete. We are interested to study this problem on a particular graph namely the generalized Petersen graph. This graph was introduced by Watkins and has a lot of interesting properties. Moreover, several graph theoretical parameters have been studied on this graph class because of its unique structure. In addition, a study of the $[1, 2]$ -total domination is also proposed at the end of this part. The last problem is a new variant called edge monitoring problem and was introduced by Dong et al. in 2008. It consists to find a set of vertices that monitors (dominates) the edge set of a graph such as a vertex monitors an edge if it forms a triangle with it i.e. it dominates both extremities of the edge. An edge can be monitored by one or more vertices. Three variants of the problem are considered in this part namely the edge monitoring, uniform edge monitoring and weighted edge monitoring. The essence of this problem lies on its combinatorial aspect and its range of applications in networks; especially wireless sensor networks. This problem is known to be NP-hard. Given the complexity of this kind of problems, we are first interested by a theoretical study: variants of the problem, bounds, characterizations, etc. We give more in depth studies of the problem for several graph classes.

Intégration des données hétérogènes issues de plusieurs fournisseurs de services géo-localisés : un cas d'utilisation sur les points d'intérêt touristique

Jury

Thomas DEVOGELE, PR, Université de Tours	Rapporteur
Ahmed LBATH, PR, Université Grenoble Alpes	Rapporteur
Philippe LAMARRE, PR, INSA Lyon	Examineur
Mauro GAIO, PR, Université de Pau et des Pays de L'Adour	Examineur
Thérèse LIBOUREL, PR, Université de Montpellier 2	Examinatrice
Christine ZANIN, MC, Université Paris Diderot	Examinatrice
Maryvonne MIQUEL, MC, INSA Lyon	Directrice de thèse
Franck FAVETTA, MC, Université Claude Bernard Lyon 1	Co-encadrant

Résumé

Les services de géolocalisation (LBS) sont impliqués dans de nombreuses applications pour fournir des informations géospatiales pertinentes basées sur une position ou une adresse géographique. La quantité de données géospatiales disponible augmente constamment et constitue des sources d'informations précieuses pour enrichir les applications LBS. Cependant, ces données géospatiales sont souvent incohérentes et contradictoires d'une source à l'autre. Aussi, pensons-nous que l'intégration de données géospatiales à partir de plusieurs sources peut améliorer la qualité de l'information offerte aux utilisateurs.

Dans cette thèse, nous nous intéresserons plus spécifiquement aux données représentant les points d'intérêt (POIs) que les touristes peuvent obtenir grâce à des applications LBS. Techniquement, un POI est représenté par une entité géospatiale qui décrit ses informations terminologiques et spatiales. La récupération, l'alignement et la fusion de ces entités géospatiales mènent à plusieurs défis. Nous nous focalisons principalement sur trois principaux défis : (i) traiter les différents schémas et structures des entités, (ii) détecter et fusionner les entités correspondantes issues de multiples sources et (iii) tenir compte de l'incertitude liée aux entités intégrées et proposer leur représentation dans les applications LBS.

Tout d'abord, nous présentons un aperçu technique qui met en évidence les méthodes utilisées par les actuels fournisseurs LBS pour partager leurs POIs ainsi que leurs limites. Ensuite, nous définissons une taxonomie de différences et d'incohérences observées entre les entités qui représentent les POIs. Cette taxonomie permet de modéliser et de comprendre comment les données peuvent différer d'une source à l'autre, ce qui nous aide à étudier comment nous devrions les intégrer. En se basant sur cette taxonomie, nous présentons PABench, un benchmark pour l'alignement des entités géospatiales. PABench peut fournir une évaluation précise des différents aspects de la qualité des approches d'alignement d'entités géospatiales et également faciliter la compréhension de leurs capacités et faiblesses quant à l'intégration géospatiale.

En ce qui concerne l'intégration des données, nous nous concentrons sur deux étapes : l'alignement d'entités et la fusion d'entités. Nous proposons l'approche Global Similarity pour l'alignement des entités géospatiales qui utilise à la fois des informations spatiales et terminologiques pour détecter les entités correspondantes. Au préalable notre approche consiste à utiliser une méthode de blocage spatial pour réduire le nombre d'entités potentiellement correspondantes. Ensuite, les entités groupées sont comparées en utilisant des mesures de similarité afin de détecter les paires correspondantes. Pour les attributs spatiaux, nous utilisons une mesure que nous avons définie et

Bilal Berjawi

Thèse

Soutenue le 01/09/2017

Établissement :
INSA Lyon

Contact :
Maryvonne.Miquel@liris.cnrs.fr

comparée à d'autres mesures existantes. Pour les attributs terminologiques, nous utilisons des mesures de similarité issues de la littérature que nous avons sélectionnées selon le type de l'attribut. Une fois les entités correspondantes détectées, un algorithme de fusion de données est mis en œuvre pour fusionner les entités correspondantes et pour estimer l'incertitude des valeurs choisies. L'incertitude sera ensuite utilisée pour informer les utilisateurs de l'exactitude des informations qu'ils reçoivent.

Enfin, nous avons étudié la visualisation d'entités fusionnées et de l'incertitude dans des cartes interactives. Nous utilisons des tests cognitifs pour déterminer les variables visuelles à utiliser et les informations à représenter directement et les informations à représenter à la demande. Nous montrons la faisabilité et l'intérêt de notre étude en développant un prototype LBS multifournisseurs et en évaluant notre proposition pour les utilisateurs potentiels.

Abstract

Location Based Services (LBS) had been involved to deliver relevant geospatial information based on a geographic position or address. The amount of geospatial data is constantly increasing, making it a valuable source of information for enriching LBS applications. However, these geospatial data are highly inconsistent and contradictory from one source to another. We assume that integrating geospatial data from several sources may improve the quality of information offered to users.

In this thesis, we specifically focus on data representing Points of Interest (POIs) that tourists can get through LBS. Technically, a POI is represented by a geospatial entity that describes the terminological and spatial information of the POI. Retrieving, matching and merging such geospatial entities lead to several challenges. We mainly focus on three main challenges including (i) dealing with different schemas and structures of entities, (ii) detecting and merging corresponding entities across multiple sources and (iii) considering the uncertainty of integrated entities and their representation in LBS applications.

First, we represent a technical overview to highlight the limitations and methods used by current LBS providers to share their POIs. Then, we define a taxonomy of observed differences and inconsistencies between the entities that represent the POIs. This taxonomy shows how data may differ from one source to another, which helps us understand how we should integrate them. Based on this taxonomy, we introduce PABench, a benchmark for geospatial entity matching. PABench can provide an accurate evaluation of the different quality aspects of geospatial entity matching approaches, and also facilitate an understanding of their weaknesses and abilities with respect to geospatial integration.

Concerning the data integration, we focus on two steps namely: entity matching and entity merging. We propose a geospatial entity matching approach namely Global Similarity that uses both spatial and terminological information to detect the corresponding entities. Our method uses a spatial blocking method to reduce the number of the potentially corresponding entities. Then, the grouped entities are compared using similarity measures in order to detect the corresponding pairs. We propose a spatial similarity measure and compare it to existing similar measures. We also compared a set of terminological similarity measures in order to select the appropriate measure to compare values of a given attribute. Once corresponding entities are detected, a data fusion algorithm is proposed to merge corresponding

entities and to estimate the uncertainty of chosen values. The uncertainty is then used to inform users about the accuracy of the information they receive.

Finally, we studied the visualization of merged entities in interactive maps. We use cognitive tests to find which visual variables to use and what information to be represented directly and what information to be represented on demand. We proved the feasibility and the benefits of our study by implementing a multi providers LBS prototype and by evaluating our proposal for potentially users.

Publications

Revue internationale avec comité de lecture

- Bilal Berjawi, Elisabeth Chesneau, Geoffrey Seccia, Fabien Duchateau, Franck Favetta, Claire Cunty, Maryvonne Miquel & Robert Laurini (2014). « Uncertainty visualization of multi-providers cartographic integration ». *Journal of Visual Languages and Computing*, vol. 25, pp. 995-1002. doi : 10.1016/j.jvlc.2014.10.033. HAL : hal-01313194.

Conférences internationales avec comité de lecture

- Léonor Ferrer Catala, Franck Favetta, Claire Cunty, Bilal Berjawi, Fabien Duchateau, Maryvonne Miquel & Robert Laurini (2016). « Visualizing Integration Uncertainty Enhances User's Choice in Multi-Providers Integrated Maps ». *International Working Conference on Advanced Visual Interfaces (AVI 2016)*, 10 juin 2016, Bari (Italie). doi : 10.1145/2909132.2926075. HAL : hal-01258627.
- Bilal Berjawi, Fabien Duchateau, Franck Favetta, Maryvonne Miquel & Robert Laurini (2015). « PABench: Designing a Taxonomy and Implementing a Benchmark for Spatial Entity Matching ». *The Seventh International Conference on Advanced Geographic Information Systems, Applications, and Services*, 27 février 2015, Lisbon (Portugal), pp. 7-16. HAL : hal-01344652.
- Anthony Morana, Thomas Morel, Bilal Berjawi & Fabien Duchateau (2014). « GeoBench: a Geospatial Integration Tool for Building a Spatial Entity Matching Benchmark (Demo) ». *International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL'2014)*, 4 novembre 2014, Dallas, Texas (États-Unis), pp. 533-536. HAL : hal-01301125.
- Bilal Berjawi, Elisabeth Chesneau, Fabien Duchateau, Franck Favetta, Claire Cunty, Maryvonne Miquel & Robert Laurini (2014). « Representing Uncertainty in Visual Integration ». *International Conference on Distributed Multimedia Systems (DMS'2014), Workshop on Visual Languages and Computing (VLC'2014)*, 27 août 2014, Wyndham Pittsburgh University Center, Pittsburgh, USA (États-Unis), pp. 365-372. HAL : hal-01301080.

Autres conférences

- Geoffrey Seccia, Claire Cunty, Elisabeth Chesneau, Bilal Berjawi & Franck Favetta (2014). « Évaluer des modes de représentation cartographique de l'incertitude ». *International Conference on Spatial Analysis and GEOmatics (MAGIS SAGEO'2014)*, 24 novembre 2014, Grenoble (France). HAL : hal-01313195.
- Bilal Berjawi (2013). « Introduction to the Integration of Location-Based Services of Several Providers ». *Bases de Données Avancées 2013 - Session Jeunes Chercheurs (BDA'2013)*, 22 octobre 2013, Nantes (France). HAL : hal-01339290.



Anytime discovery of a diverse set of patterns with Monte Carlo tree search

Jury

Toon CALDERS, PR, Univesiteit Antwerpe
Tristan CAZENAVE, PR, Université Paris-Dauphine
Sihem AMER-YAHIA, PR, CNRS
Moustafa BENSAFI, DR, CNRS
Peter FLACH, PR, University Bristol
Katharina MORIK, PR, Technische Universität Dortmund
Jean-François BOULICAUT, PR, INSA Lyon
Mehdi KAYTOUE, MC, INSA Lyon

Rapporteur
Rapporteur
Examinatrice
Examineur
Examineur
Examinatrice
Directeur de thèse
Co-directeur

Résumé

La découverte de motifs qui caractérisent fortement une classe vis à vis d'une autre reste encore un problème difficile en fouille de données. La découverte de sous-groupes (Subgroup Discovery, SD) est une approche formelle de fouille de motifs qui permet la construction de classifieurs intelligibles mais surtout d'émettre des hypothèses sur les données. Cependant, cette approche fait encore face à deux problèmes majeurs : (i) comment définir des mesures de qualité appropriées pour caractériser l'intérêt d'un motif et (ii) comment sélectionner une méthode heuristique adaptée lorsqu'une énumération exhaustive de l'espace de recherche n'est pas réalisable. Le premier problème a été résolu par la fouille de modèles exceptionnels (Exceptional Model Mining, EMM) qui permet l'extraction de motifs couvrant des objets de la base de données pour lesquels le modèle induit sur les attributs de classe est significativement différent du modèle induit par l'ensemble des objets du jeu de données. Le second problème a été étudié en SD et EMM principalement avec la mise en place de méthodes heuristiques de type recherche en faisceau (beam-search) ou avec des algorithmes génétiques qui permettent la découverte de motifs non redondants, diversifiés et de bonne qualité. Dans cette thèse, nous soutenons que la nature gloutonne des méthodes d'énumération précédentes génère cependant des ensembles de motifs manquant de diversité. Nous définissons formellement la fouille de données comme un jeu que nous résolvons par l'utilisation de la recherche arborescente de Monte Carlo (Monte Carlo Tree Search, MCTS), une technique récente principalement utilisée pour la résolution de jeux et de problèmes de planning en intelligence artificielle. Contrairement aux méthodes traditionnelles d'échantillonnage, MCTS donne la possibilité d'obtenir une solution à tout instant sans qu'aucune hypothèse ne soit faite que ce soit sur la mesure de qualité ou sur les données. Cette méthode d'énumération converge vers une approche exhaustive si les budgets temps et mémoire disponibles sont suffisants. Le compromis entre l'exploration et l'exploitation que propose cette approche permet une augmentation significative de la diversité dans l'ensemble des motifs calculés. Nous montrons que la recherche arborescente de Monte Carlo appliquée à la fouille de motifs permet de trouver rapidement un ensemble de motifs diversifiés et de bonne qualité à l'aide d'expérimentations sur des jeux de données de référence et sur un jeu de données réel traitant de l'olfaction. Nous proposons et validons également une nouvelle mesure de qualité spécialement conçue pour des jeux de données multi labels présentant une grande variance de fréquences des labels.

Abstract

The discovery of patterns that strongly distinguish one class label from another is still a challenging data-mining task. Subgroup Discovery (SD) is a

Guillaume Bosc

Thèse

Soutenue le 11/09/2017

Établissement :
INSA Lyon

Contact :

Jean-François-Boulicaut@liris.cnrs.fr

formal pattern mining framework that enables the construction of intelligible classifiers, and, most importantly, to elicit interesting hypotheses from the data. However, SD still faces two major issues: (i) how to define appropriate quality measures to characterize the interestingness of a pattern; (ii) how to select an accurate heuristic search technique when exhaustive enumeration of the pattern space is unfeasible. The first issue has been tackled by Exceptional Model Mining (EMM) for discovering patterns that cover tuples that locally induce a model substantially different from the model of the whole dataset. The second issue has been studied in SD and EMM mainly with the use of beam-search strategies and genetic algorithms for discovering a pattern set that is non-redundant, diverse and of high quality. In this thesis, we argue that the greedy nature of most such previous approaches produces pattern sets that lack diversity. Consequently, we formally define pattern mining as a game and solve it with Monte Carlo Tree Search (MCTS), a recent technique mainly used for games and planning problems in artificial intelligence. Contrary to traditional sampling methods, MCTS leads to an any-time pattern mining approach without assumptions on either the quality measure or the data. It converges to an exhaustive search if given enough time and memory. The exploration/exploitation trade-off allows the diversity of the result set to be improved considerably compared to existing heuristics. We show that MCTS quickly finds a diverse pattern set of high quality in our application in neurosciences. We also propose and validate a new quality measure especially tuned for imbalanced multi-label data.

Publications

Conférences internationales avec comité de lecture

- Guillaume Bosc, Jean-François Boulicaut, Chedy Raïssi & Mehdi Kaytoue (2017). « Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search ». *Data Mining and Knowledge Discovery*. doi : 10.1007/s10618-017-0547-5. HAL : hal-01662857.
- Guillaume Bosc, Philip Tan, Jean-François Boulicaut, Chedy Raïssi & Mehdi Kaytoue (2017). « A Pattern Mining Approach to Study Strategy Balance in RTS Games ». *IEEE Transactions on Computational Intelligence and AI in games*, vol. 9, n°2, pp. 123-132. doi : 10.1109/TCIAIG.2015.2511819. HAL : hal-01252728.

Conférences nationales avec comité de lecture

- Victor Codocedo, Guillaume Bosc, Mehdi Kaytoue, Jean-François Boulicaut & Amedeo Napoli (2017). « A Proposition for Sequence Mining Using Pattern Structures ». *ICFCA 2017 - 14th International Conference on Formal Concept Analysis*, 16 juin 2017, Rennes (France), pp. 106-121. doi : 10.1007/978-3-319-59271-8_7. HAL : hal-01549107.
- Guillaume Bosc, Marc Plantevit, Jean-François Boulicaut, Moustafa Bensafi & Mehdi Kaytoue (2016). « h(odor): Interactive Discovery of Hypotheses on the Structure-Odor Relationship in Neuroscience ». *ECML/PKDD 2016 (Demo)*, 23 septembre 2016, Riva del Garda (Italie). HAL : hal-01346679.
- Guillaume Bosc, Jérôme Golebiowski, Moustafa Bensafi, Céline Robardet, Marc Plantevit, Jean-François Boulicaut & Mehdi Kaytoue (2016). « Local subgroup discovery for eliciting and understanding new structure-odor relationships ». *Discovery Science: 19th International Conference, DS 2016*, 21 octobre 2016, Bari (Italie), pp. 19-34. doi : 10.1007/978-3-319-46307-0_2. HAL : hal-01346660.

- Guillaume Bosc, Mehdi Kaytoue, Chedy Raïssi, Jean-François Boulicaut & Philip Tan (2014). « Mining Balanced Sequential Patterns in RTS Games 1 ». ECAI 2014 - 21st European Conference on Artificial Intelligence, 22 août 2014, Prague (République Tchèque). doi : 10.3233/978-1-61499-419-0-975. HAL : hal-011100933.

Conférences nationales avec comité de lecture

- Guillaume Bosc, Jean-François Boulicaut, Chedy Raïssi & Mehdi Kaytoue (2017). « Découverte de sous-groupes avec les arbres de recherche de Monte Carlo ». Extraction et Gestion de Connaissances EGC 2017, 27 janvier 2017, Grenoble (France). HAL : hal-01433054.
- Guillaume Bosc, Mehdi Kaytoue, Marc Plantevit, Fabien De Marchi, Moustafa Bensafi & Jean-François Boulicaut (2015). « Vers la découverte de modèles exceptionnels locaux : des règles descriptives liant les molécules à leurs odeurs ». 15e Journées Internationales Francophones Extraction et Gestion des Connaissances (EGC 2015), 30 janvier 2015, Luxembourg (Luxembourg). HAL : hal-01346739.
- Guillaume Bosc, Mehdi Kaytoue, Chedy Raïssi & Jean-François Boulicaut (2014). « Fouille de motifs séquentiels pour l'élicitation de stratégies à partir de traces d'interactions entre agents en compétition ». Extraction et Gestion de Connaissances EGC 2014, 28 janvier 2014, Rennes (France), pp. 359-370. HAL : hal-01270794.

Autres

- Guillaume Bosc, Jean-François Boulicaut, Chedy Raïssi & Mehdi Kaytoue (2016). « Anytime Discovery of a Diverse Set of Patterns with Monte Carlo Tree Search ». ArXiv : 1609.08827v2. HAL : hal-01418663.



Compression progressive de maillages surfaciques texturés

Jury

ALLIEZ Pierre, DR, INRIA Sophia-Antipolis
BONNEAU George-Pierre, PR, Université Grenoble Alpes
BECHMANN Dominique, PR, Université de Strasbourg
HUDELOT Céline, MC, École Centrale Paris
LAVOUÉ Guillaume, MC, INSA Lyon
VIDAL Vincent, MC, Université Claude Bernard Lyon
DUPONT Florent, PR, Université Claude Bernard Lyon

Rapporteur
Rapporteur
Examinatrice
Examinatrice
Directeur de thèse
Co-encadrant
Examineur

Résumé

Depuis plusieurs années, les modèles 3D deviennent de plus en plus détaillés. Cela augmente considérablement le volume de données les décrivant. Cependant, dans un même temps, un nombre croissant d'applications sont contraintes en mémoire et/ou en vitesse (visualisation sur périphériques mobiles, jeux vidéos, etc.). Dans un contexte Web, ces difficultés sont encore plus présentes. Cette situation peut entraîner des incompatibilités, des latences de transmission ou d'affichage qui sont souvent problématiques.

La compression progressive de ces modèles est une des solutions envisageables. Le but étant de compresser les informations (géométrie, connectivité et attributs associés) de façon à pouvoir reconstruire progressivement le maillage. À la différence d'une compression dite single-rate, la compression progressive propose très rapidement un aperçu fidèle du modèle 3D pour ensuite le raffiner jusqu'à retrouver le maillage complet. Ceci permet un meilleur confort pour l'utilisateur et une adaptation de la quantité d'éléments à visualiser ou à traiter en fonction des capacités du périphérique de réception.

Généralement, les approches existantes pour la compression progressive se focalisent sur le traitement de maillages 2-variétés triangulaires. Très peu de méthodes sont capables de compresser progressivement des maillages surfaciques non-variétés et, à notre connaissance, aucune ne permet de compresser génériquement un maillage surfacique quel que soit son type (i.e. non-variété et polygonal). Pour supprimer ces limitations, nous présentons une méthode de compression progressive générique permettant de traiter l'ensemble des maillages surfaciques (non-variétés et polygonaux). De plus, notre approche tient compte de l'attribut de texture potentiellement associé à ces maillages, en gérant correctement les coutures éventuelles.

Pour ce faire, nous décimons progressivement le maillage à l'aide d'un nouvel opérateur générique de simplification. Cette décimation est guidée par une métrique locale qui a pour but de préserver la géométrie et la paramétrisation de la texture. Durant cette simplification, nous encodons progressivement les informations nécessaires à la reconstruction. Afin d'améliorer le taux de compression, nous mettons en oeuvre certains procédés de réduction de l'entropie, ainsi que des dispositifs de prédiction basés sur la géométrie pour l'encodage de la connectivité et des coordonnées de texture. Pour finir, l'image de texture est compressée progressivement puis multiplexée avec les données relatives au maillage. Ce multiplexage est réalisé à l'aide d'une métrique perceptuelle afin d'obtenir le meilleur rapport débit-distorsion possible lors de la décompression.

Florian Caillaud

Thèse

Soutenue le 17/01/2017

Établissement :
INSA Lyon

Contact :
Guillaume.Lavoue@liris.cnrs.fr

Abstract

Since several years, 3D models become more and more detailed. This increases substantially the amount of data needed to describe them. However, in the same time, a rising number of applications are constrained in memory and/or in speed (mobile device visualization, video games, etc.). These difficulties are even more visible within a Web context. This situation can lead to incompatibilities, latency in transmission or rendering, which is generally an issue.

The progressive compression of these models is a possible solution. The goal is to compress the information (geometry, connectivity and associated attributes) in order to allow a progressive reconstruction of the mesh. Contrary to a single-rate compression, progressive compression quickly proposes a faithful draft of the 3D model and, then, refines it until the complete mesh is recovered. This allows a better comfort for the user and a real adaptation of the rendered element number in adequacy with the terminal device properties.

The existing approaches for progressive compression mainly focus on triangular 2- manifold meshes. Very few methods are able to compress progressively non-manifold surface meshes and, to our knowledge, none can deal with every surface meshes (i.e. non-manifold and polygonal), in a generic way. So as to suppress these limitations, we present a new generic progressive method allowing the compression of polygonal non-manifold surface meshes. Moreover, our approach takes care of the texture attribute, possibly associated to these meshes, by handling properly potential texture seams.

For that purpose, we progressively decimate the mesh using a new generic simplification operator. This decimation is driven by a local metric which aims to preserve both the geometry and the texture parametrisation. During the simplification, we progressively encode the necessary information for the reconstruction. In order to improve the compression rate, we propose several entropy reduction mechanisms, as well as geometry based prediction strategies for the connectivity and UV coordinates encoding. Finally, the texture map is progressively compressed then multiplexed with mesh data. This multiplexing is achieved using a perceptual metric to provide the best rate-distortion ratio as possible during the decompression.

Publications

Revue internationale avec comité de lecture

- Florian Caillaud, Vincent Vidal, Florent Dupont & Guillaume Lavoué (2016). « Progressive Compression of Arbitrary Textured Meshes ». Computer Graphics Forum, vol. 35, n°7, pp. 475-484. doi : 10.1111/cgf.13044. HAL : hal-01376105.

Conférences internationales avec comité de lecture

- Florian Caillaud, Vincent Vidal, Florent Dupont & Guillaume Lavoué (2015). « Progressive compression of generic surface meshes ». Computer Graphics International 2015 (CGI'15), 26 juin 2015, Strasbourg (France). HAL : hal-01191857.

- F. Caillaud, X. Faure, F. Zara, F. Jaillet & J. -M. Moreau (2013). « Multi-criteria adaptation of physical simulations ». Eurographics Workshop on Virtual Reality Interaction and Physical Simulation (VRIPHYS 2013), 29 novembre 2013, Lille (France), pp. 1-4. HAL : hal-00949836.

Conférences nationales avec comité de lecture

- Florian Caillaud, Guillaume Lavoué, Vincent Vidal & Florent Dupont (2014). « Compression Progressive Générique de Maillages Surfaiques ». Reims Image, 26 novembre 2014, Reims (France), pp. 1-8. HAL : hal-01313187.

Autres

- Guillaume Lavoué, Laurent Chevalier, Florian Caillaud & Florent Dupont (2016). « Progressive streaming of textured 3D models in a web browser ». ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, 27 février 2016, Redmond, Washington (États-Unis). HAL : hal-01648970.



Mécanismes de négociation distribuée pour la Gestion intelligente de l'énergie

Romain Cailliere

Jury

René MANDIAU, PR, Université de Valenciennes et du Hainaut Cambrésis
Rapporteur

Leïla MERGHEM, MC, Université de Technologie de Troyes Rapporteuse

Onn SHEHORY, PR, Bar Ilan University Examineur

Elise BONZON, MC, Université Paris Descartes - Paris 5 Examinatrice

Samir AKNINE, PR, Université Claude Bernard Lyon 1 Directeur de thèse

Antoine NONGAILLARD, MC, Université des Sciences et Technologies de Lille
Co-directeur

Résumé

Les besoins, toujours plus grands, en énergie et la pollution de la planète, due à l'utilisation d'énergies polluantes non-renouvelables, obligent à concevoir de nouveaux modèles énergétiques durables et fiables. Ces nouveaux modèles se fondent, aujourd'hui, sur une intégration massive des énergies renouvelables dans le réseau électrique. Le problème des énergies renouvelables est leur caractère intermittent, dépendant des conditions météorologiques, la plupart du temps. L'arrivée des nouvelles technologies de l'information et de la communication permet l'intégration d'une couche informationnelle au réseau énergétique lui permettant d'être intelligent et d'entrevoir la possibilité d'une gestion distribuée des énergies renouvelables. Ces énergies étant principalement décentralisées, contrairement aux imposantes centrales nucléaires, au charbon, au gaz et au fioul, sont produites directement chez le consommateur. Le consommateur devient alors un prosumer capable de répondre à ses besoins énergétiques, voire même d'agir comme un producteur s'il produit plus d'énergie qu'il n'en consomme. Mais l'arrivée d'une pléthore de petits acteurs capables d'acheter et de vendre de l'énergie, en temps réel, dans un marché comprenant les puissants acteurs traditionnels du marché peut être une source de volatilité pour les prix de l'énergie. Des variations importantes des prix peuvent conduire à des situations néfastes en déstabilisant le réseau. Pour faire face à ce problème, nous avons développé un premier mécanisme de négociation automatique, sur trois échelles de temps, qui impose des contraintes sur la demande ainsi que sur les prix pour garantir leur stabilité. Ce mécanisme s'appuie sur des entités représentatives (producteurs, prosumers et agrégateurs) pour gérer l'offre et la demande sans toutefois prendre en compte l'impact sur le réseau des contrats négociés entre ces entités. Le second mécanisme, fondé sur la technologie Blockchain, permet des négociations bilatérales décentralisées et prend en compte les impacts physiques de chaque échange d'énergie entre prosumers, garantissant ainsi l'intégrité du réseau. Le mécanisme ainsi proposé se soustrait à une gestion de la stabilité du réseau par des tiers.

Publications

Conférences internationales avec comité de lecture

- Romain Caillière, Samir Aknine & Antoine Nongillard (2016). « Multi-agent mechanism for efficient cooperative use of energy ». AAMAS, Singapour (Singapour), 1365—1366. HAL : hal-01261596.
- Arame Ndeye Diago, Samir Aknine, Onn Shehory, Souhila Arib, Romain Caillière & Mbaye Sène (2016). « Decentralized and fair multilateral negotiation ». ICTAI, California (États-Unis), 149—156. HAL : hal-01355874.

Thèse

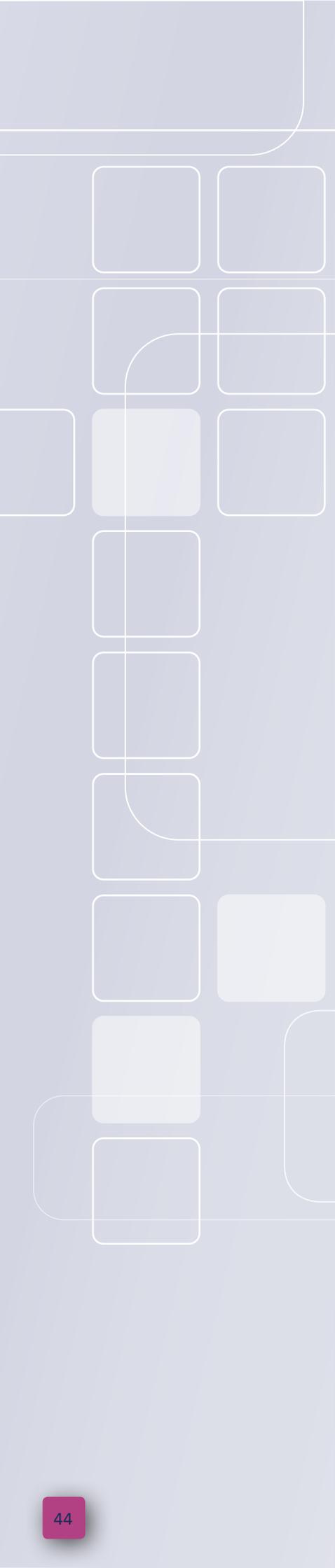
Soutenue le 05/09/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Samir.Aknine@liris.cnrs.fr

- 
- Romain Caillière, Souhila Arib, Samir Aknine & Chantal Berdier (2014). « Bi-level token ring for multilateral negotiations ». International Conference on Intelligent Agent Technology, 11 août 2014, WARSAW (Pologne), pp. 9-16. doi : 10.1109/WI-IAT.2014.144. HAL : hal-01301061.

Détection d'objets en mouvement à l'aide d'une caméra mobile

Jury

Pascal BERTOLINO, MC, Université Grenoble Alpes
Frédéric CORDIER, MC, Université de Haute Alsace
Albert DIPANDA, PR, Université de Bourgogne
Karine ZEITOUNI, PR, Université de Versailles St Quentin
Sylvie CALABRETTO, PR, INSA Lyon
Erwan GUILLOU, MC, Lyon 1
Saida BOUAKAZ, PR, Lyon 1

Rapporteur
Rapporteur
Examineur
Examinatrice
Examinatrice
Co-directeur
Directrice de thèse

Résumé

La détection d'objets mobiles dans des flux vidéo est une étape essentielle pour de nombreux algorithmes de vision par ordinateur. Cette tâche se complexifie lorsque la caméra utilisée est en mouvement. En effet, l'environnement capté par ce type de caméra apparaît en mouvement et il devient plus difficile de distinguer les objets qui effectuent réellement un mouvement de ceux qui constituent la partie statique de la scène.

Dans cette thèse, nous apportons des contributions au problème de détection d'objets mobiles dans le flux vidéo d'une caméra mobile. L'idée principale qui nous permet de distinguer les éléments mobiles de ceux qui sont statiques repose sur un calcul de distance dans l'espace 3D. Les positions 3D de caractéristiques extraites des images sont estimées par triangulation puis leurs mouvements 3D sont analysés pour réaliser un étiquetage épars statique/mobile de ces points. Afin de rendre la détection robuste au bruit, l'analyse des mouvements 3D des caractéristiques est comparée à d'autres points précédemment estimés statiques. Une mesure de confiance, mise à jour au cours du temps, est utilisée pour déterminer l'étiquette à attribuer à chacun des points.

Nos contributions ont été appliquées à des jeux de données virtuelles (issus du projet Previz) et réelles (reconnus dans la communauté) et les comparaisons ont été réalisées avec l'état de l'art. Les résultats obtenus montrent que la contrainte 3D proposée dans cette thèse, couplée à une analyse statistique et temporelle des mouvements, permet de détecter des éléments mobiles dans le flux vidéo d'une caméra en mouvement et ce même dans des cas complexes où les mouvements apparents de la scène ne sont pas uniformes.

Abstract

Moving objects detection in video streams is a commonly used technique in many computer vision algorithms. The detection becomes more complex when the camera is moving. The environment observed by this type of camera appeared moving and it is more difficult to distinguish the objects which are in movement from the others that composed the static part of the scene.

In this thesis we propose contributions for the detection of moving objects in the video stream of a moving camera. The main idea to differentiate between moving and static objects based on 3D distances. 3D positions of feature points extracted from images are estimated by triangulation and then their 3D motions are analyzed in order to provide a sparse static/moving labeling. To provide a more robust detection, the analysis of the 3D motions is compared to those of feature points previously estimated static. A confidence value updated over time is used to decide on labels to attribute to each point.

Marie-Neige Chapel

Thèse

Soutenue le 22/09/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Saida.Bouakaz@liris.cnrs.fr

We make experiments on virtual (from the Previz project) and real datasets (known by the community) and we compare the results with the state of the art. The results show that our 3D constraint coupled with a statistical and temporal analysis of motions allow to detect moving elements in the video stream of a moving camera even in complex cases where apparent motions of the scene are not similars.

Publications

Conférences internationales avec comité de lecture

- Marie-Neige Chapel, Erwan Guillou & Saida Bouakaz (2017). « Coupled 2D and 3D Analysis for Moving Objects Detection with a Moving Camera ». 12th International Conference on Computer Vision Theory and Applications, 1 mars 2017, Porto (Portugal), pp. 236-245. doi : 10.5220/0006132002360245. HAL : hal-01488610.

Autres

- Marie-Neige Chapel & Erwan Guillou (2014). « Segmentation de séquence vidéo en plans cohérents pour compositing 2D ». Journées de l'Association Française d'Informatique Graphique (AFIG), 28 novembre 2014, Reims (France). HAL : hal-01128078.

Approche déclarative pour le stockage à long terme de données capteur

Jury

Nicole BIDOIT, PR, Université Paris Sud 11
Jef WIJSEN, PR, Université de Mons
Amel BOUZEGHOUB, PR, Telecom SudParis
Lotfi LAKHAL, PR, Aix-Marseille Université
Jean-Marc PETIT, PR, INSA Lyon
Yann GRIPAY, MC, INSA Lyon
Nicolas FOURTY, MC, Université Pierre Mendès
Sadok BEN YAHIA, PR, Faculté des Sciences de Tunis

Rapporteuse
Rapporteur
Examinatrice
Examineur
Directeur de thèse
Co-directeur
Co-encadrant
Invité

Manel Charfi

Résumé

De nos jours, on a de plus en plus de capteurs qui ont tendance à apporter confort et facilité dans notre vie quotidienne. Ces capteurs sont faciles à déployer et à intégrer dans une variété d'applications (monitoring de bâtiments intelligents, aide à la personne,...). Ces milliers (voire millions) de capteurs sont de plus en plus envahissants et génèrent sans arrêt des masses énormes de données qu'on doit stocker et gérer pour le bon fonctionnement des applications qui en dépendent.

A chaque fois qu'un capteur génère une donnée, deux dimensions sont d'un intérêt particulier : la dimension temporelle et la dimension spatiale. Ces deux dimensions permettent d'identifier l'instant de réception et la source émettrice de chaque donnée. Chaque dimension peut se voir associée à une hiérarchie de granularités qui peut varier selon le contexte d'application.

Dans cette thèse, nous nous concentrons sur les applications nécessitant une conservation à long terme des données issues des flux de données capteurs. Notre approche vise à contrôler le stockage des données capteurs en ne gardant que les données jugées pertinentes selon la spécification des granularités spatio-temporelles représentatives des besoins applicatifs, afin d'améliorer l'efficacité de certaines requêtes. Notre idée clé consiste à emprunter l'approche déclarative développée pour la conception de bases de données à partir de contraintes et d'étendre les dépendances fonctionnelles avec des composantes spatiales et temporelles afin de revoir le processus classique de normalisation de schéma de base de données.

Étant donné des flux de données capteurs, nous considérons à la fois les hiérarchies de granularités spatio-temporelles et les Dépendances Fonctionnelles SpatioTemporelles (DFSTs) comme objets de premier ordre pour concevoir des bases de données de capteurs compatibles avec n'importe quel SGBDR.

Nous avons implémenté un prototype de cette architecture qui traite à la fois la conception de la base de données ainsi que le chargement des données. Nous avons mené des expériences avec des flux de données synthétiques et réels provenant de bâtiments intelligents. Nous avons comparé notre solution avec la solution de base et nous avons obtenu des résultats prometteurs en termes de performance de requêtes et d'utilisation de la mémoire. Nous avons également étudié le compromis entre la réduction des données et l'approximation des données.

Thèse

Soutenue le 21/09/2017

Établissement :
INSA Lyon

Contact :
Jean-Marc.Petit@liris.cnrs.fr

Abstract

Nowadays, sensors are cheap, easy to deploy and immediate to integrate into applications. These thousands of sensors are increasingly invasive and are constantly generating enormous amounts of data that must be stored and managed for the proper functioning of the applications depending on them. Sensor data, in addition of being of major interest in real-time applications, e.g. building control, health supervision..., are also important for long-term reporting applications, e.g. reporting, statistics, research data...

Whenever a sensor produces data, two dimensions are of particular interest: the temporal dimension to stamp the produced value at a particular time and the spatial dimension to identify the location of the sensor. Both dimensions have different granularities that can be organized into hierarchies specific to the concerned context application.

In this PhD thesis, we focus on applications that require long-term storage of sensor data issued from sensor data streams. Since huge amount of sensor data can be generated, our main goal is to select only relevant data to be saved for further usage, in particular long-term query facilities. More precisely, our aim is to develop an approach that controls the storage of sensor data by keeping only the data considered as relevant according to the spatial and temporal granularities representative of the application requirements. In such cases, approximating data in order to reduce the quantity of stored values enhances the efficiency of those queries. Our key idea is to borrow the declarative approach developed in the seventies for database design from constraints and to extend functional dependencies with spatial and temporal components in order to revisit the classical database schema normalization process.

Given sensor data streams, we consider both spatio-temporal granularity hierarchies and Spatio-Temporal Functional Dependencies (STFDs) as first class-citizens for designing sensor databases on top of any RDBMS. We propose a specific axiomatisation of STFDs and the associated attribute closure algorithm, leading to a new normalization algorithm.

We have implemented a prototype of this architecture to deal with both database design and data loading. We conducted experiments with synthetic and real-life data streams from intelligent buildings.

Publications

Conférences internationales avec comité de lecture

- Manel Charfi, Yann Gripay & Jean-Marc Petit (2017). « Spatio-Temporal Functional Dependencies for Sensor Data Streams ». SSTD'17, 21 août 2017, Arlington, Virginia (États-Unis), pp. 182-199. doi : 10.1007/978-3-319-64367-0_10. HAL : hal-01527525.
- Manel Charfi, Yann Gripay & Jean-Marc Petit (2016). « Optimization of a Class of Temporal Queries ». IDEAS '16, 13 juillet 2016, Montreal, QC (Canada). doi : 10.1145/2938503.2938514. HAL : hal-01520719.

Conférences nationales avec comité de lecture

- Manel Charfi, Yann Gripay & Jean-Marc Petit (2016). « Conception d'une base de données capteurs à l'aide de contraintes spatio-temporelles ». BDA 2016, Poitiers (France). HAL : hal-01520716.
- Manel Charfi, Yann Gripay, Nicolas Fourty, Denis Genon-Catalot & Jean-Marc Petit (2014). « Approche déclarative pour le monitoring des bâtiments intelligents ». Journées Nationales des Communications Terrestres (JNCT 2014), 22 mai 2014, Toulouse-Blagnac (France), pp. 123-131. HAL : hal-01301107.
- Manel Charfi (2014). « Optimisation sémantique des requêtes continues Application aux bâtiments intelligents ». BDA 2014 : Gestion de données - principes, technologies et applications, 17 octobre 2014, Autrans (France), pp. 26-27. HAL : hal-01169952.

Autres

- Manel Charfi (2014). « Optimisation sémantique des requêtes continues : Application aux bâtiments intelligents ». BDA 2014, Bases de Données Avancées 2014, Session «Jeunes chercheurs», 14 octobre 2014, Grenoble-Autrans (France). HAL : hal-01301108.



Acquisition de connaissances pour des EIAH destinés à enseigner des méthodes de résolution de problèmes

Awa Diattara

Jury

Marie-Hélène ABEL, PR, UTC, Compiègne
Serge GARLATTI, PR, IMT Atlantique, Brest
Hamid CHAACHOUA, PR, Université Grenoble Alpes
Vanda LUENGO, PR, UPMC, Paris
Nathalie GUIN, MC, UCBL, Lyon

Rapporteur
Rapporteur
Examineur
Directrice de thèse
Co-directrice de thèse

Résumé

Cette thèse aborde la problématique de l'acquisition des connaissances dans le cadre de la conception des Environnements Informatiques pour l'Apprentissage Humain (EIAH). Dans le contexte spécifique de ce travail, nous nous intéressons à des EIAH destinés à enseigner des méthodes de résolution de problèmes. De telles méthodes permettent, dans un domaine donné, de reconnaître la classe d'un problème et d'être capable de savoir quelle technique de résolution appliquer pour le résoudre. Le coût de conception de ces EIAH est cependant très élevé, en particulier du fait de l'élicitation des connaissances qui nécessite non seulement une expertise dans le domaine concerné mais également en programmation.

Afin de réduire le coût de conception de ces EIAH et permettre à des auteurs (enseignants, professionnels ou pédagogues plus experts) de pouvoir éliciter sans programmer, les connaissances nécessaires à l'EIAH pour résoudre des problèmes, nous proposons un processus d'acquisition interactive de ces connaissances. Ce processus est mis en œuvre à travers la conception d'un outil auteur : AMBRE-KB. Pour ce faire, nous avons d'abord proposé des méta-modèles qui permettent de décrire la forme des connaissances à acquérir. Ces connaissances ne sont pas celles d'un expert dans un domaine, mais les connaissances telles qu'on voudrait qu'elles fonctionnent chez l'apprenant à l'issue de l'apprentissage. En s'appuyant sur ces méta-modèles, nous avons ensuite proposé un processus d'acquisition de ces connaissances qui assiste l'auteur lors de l'élicitation des connaissances nécessaires, en lui permettant de construire un modèle de connaissances spécifique à un domaine.

Deux expérimentations ont été menées afin de mettre à l'essai AMBRE-KB. Les résultats de la première expérimentation - réalisée sur trois domaines - ont conduit à faire évoluer les fonctionnalités et les interfaces de AMBRE-KB afin d'améliorer l'interactivité du système, mais également permettre à l'auteur d'avoir un feedback sur le processus d'élicitation. Cette nouvelle version de AMBRE-KB a permis de mener une seconde expérimentation réalisée sur quatre domaines. Celle-ci a permis de mettre en évidence les points forts et les points faibles de AMBRE-KB :

Pour un domaine donné, AMBRE-KB permet de générer les modèles de connaissances permettant au résolveur de résoudre les problèmes relevant de la méthode à enseigner

- Les interfaces proposées sont faciles à utiliser dans l'ensemble
- Les auteurs ont apprécié l'assistance fournie, en particulier le feedback

Une amélioration de certaines interfaces et un feedback plus poussé faciliteraient davantage le processus d'élicitation.

Thèse

Soutenue le 20/10/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Nathalie.Guin@liris.cnrs.fr

Abstract

The general issue we addressed in this thesis is the challenge of knowledge elicitation in Intelligent Tutoring Systems (ITS). In the context of this work, we are interested in ITS teaching problem solving methods. Teaching methods for solving problems consists in teaching students how to think about the problem before starting its resolution. In a given field, such a method is based on a categorization of problems. Knowing to recognize the class of a problem enables students to choose the resolution technique associated with this class. However, designing such ITS is tedious and costly, and specially require expertise in the application domain and in programming.

In order to reduce the design cost of these ITS and to enable an author (for example a teacher) to be able to elicit knowledge needed without programming, we propose an interactive knowledge elicitation process. This process is implemented through the design of an authoring tool: AMBRE-KB. For that, we first propose meta-models for the knowledge to be acquired. This knowledge is not an expert knowledge, but knowledge such as we would want that it works at the end of the learning. Next, we propose a knowledge acquisition process based on these meta-models, which enable the author to be assisted in the elicitation process enabling him/her to build specific knowledge models for a given domain.

Two experiments were conducted to evaluate the knowledge acquisition process and its implementation in the AMBRE-KB tool. The results of the first experiment, conducted in three domains, led to modify the features and the interfaces of AMBRE-KB, so that to improve the interactivity of the tool, and to enable authors to have feedback about the knowledge elicitation. The second experiment, based on the new version of AMBRE-KB, was conducted in four domains. The results of this experiment enable to highlight the strengths and the limits of AMBRE-KB:

For a given domain, AMBRE-KB can produce the knowledge models needed by the solver to solve problems about the method to be taught

- On the whole, interfaces are easy to use
- The assistance provided is useful and particularly the feedback

An improvement of the interfaces and of the interactivity of the system would facilitate even more the elicitation process

Publications

Conférences internationales avec comité de lecture

- Awa Diattara, Nathalie Guin, Vanda Luengo & Amélie Cordier (2017). « An Authoring Tool to Elicit Knowledge to be Taught without Programming ». International Conference on Computer Supported Education, 23 avril 2017, Porto (Portugal), pp. 82-91. doi : 10.5220/0006315100820091. HAL : hal-01588822.
- Awa Diattara, Nathalie Guin, Vanda Luengo & Amélie Cordier (2016). « Towards an Authoring Tool to Acquire Knowledge for ITS Teaching Problem Solving Methods ». European Conference on Technology Enhanced Learning (EC-TEL 2016), 16 septembre 2016, Lyon (France), pp. 575-578. doi : 10.1007/978-3-319-45153-4_66. HAL : hal-01357997.

Conférences nationales avec comité de lecture

- Awa Diattara (2016). « Vers un outil auteur pour des EIAH destinés à l'apprentissage de méthodes de résolution de problèmes : acquisition des connaissances sur la méthode à enseigner ». Rencontres Jeunes Chercheurs en EIAH (RJC-EIAH 2016), 17 août 2016, Montpellier (France). HAL : hal-01357989.
- Awa Diattara, Nathalie Guin, Marie Lefèvre, Bryan Kong Win Chang & Vanda Luengo (2014). « Vers un outil auteur pour des EIAH destinés à l'apprentissage de méthodes : formalisation des connaissances à acquérir ». TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION POUR L'ENSEIGNEMENT, 20 novembre 2014, Béziers (France). HAL : hal-01283735.



Estimation de pose humaine et reconnaissance d'actions par un système multi-robots

Jury

David FOFI, PR, Université de Bourgogne
Nicole VINCENT, PR, Université Paris Descartes
Denis PELLERIN, PR, Institut polytechnique de Grenoble
Christophe DUCOTTET, PR, Université Jean Monnet
Céline TEULIÈRE, MC, Institut Pascal
Atilla BASKURT, PR, INSA Lyon
Christian WOLF, MC, INSA Lyon
Gonen EREN, MC, Université Galatasaray

Rapporteur
Rapporteur
Examineur
Examineur
Examinatrice
Directeur de thèse
Co-directeur
Co-directeur

Emre Dogan

Résumé

L'estimation de la pose humaine et la reconnaissance des activités humaines sont des étapes importantes dans de nombreuses applications comme la robotique, la surveillance et la sécurité, etc. Actuellement abordées dans le domaine, ces tâches ne sont toujours pas résolues dans des environnements non-coopératifs particulièrement. Ces tâches admettent de divers défis comme l'occlusion, les variations des vêtements, etc. Les méthodes qui exploitent des images de profondeur ont l'avantage concernant les défis liés à l'arrière-plan et à l'apparence, pourtant, l'application est limitée pour des raisons matérielles. Dans un premier temps, nous nous sommes concentrés sur la reconnaissance des actions complexes depuis des vidéos. Pour ceci, nous avons introduit une représentation spatio-temporelle indépendante du point de vue. Plus précisément, nous avons capturé le mouvement de la personne en utilisant un capteur de profondeur et l'avons encodé en 3D pour le représenter. Un descripteur 3D a ensuite été utilisé pour la classification des séquences avec la méthodologie bag-of-words. Pour la deuxième partie, notre objectif était l'estimation de pose articulée, qui est souvent une étape intermédiaire pour la reconnaissance de l'activité. Notre motivation était d'incorporer des informations à partir de capteurs multiples et de les fusionner pour surmonter le problème de l'autoocclusion. Ainsi, nous avons proposé un modèle de flexible mixtures-of-parts multivues inspiré par la méthodologie classique de structure pictural. Nous avons démontré que les contraintes géométriques et les paramètres de cohérence d'apparence sont efficaces pour renforcer la cohérence entre les points de vue, aussi que les paramètres classiques. Finalement, nous avons évalué ces nouvelles méthodes sur des datasets publics, qui vérifie que l'utilisation de représentations indépendantes de la vue et l'intégration d'informations à partir de points de vue multiples améliore la performance pour les tâches ciblées dans le cadre de cette manuscrit.

Publications

Revue internationale avec comité de lecture

- Emre Dogan, Gonen Eren, Christian Wolf, Eric Lombardi & Atilla Baskurt (2018). « Multi-view pose estimation with mixtures-of-parts and adaptive viewpoint selection ». IET Computer Vision. HAL : hal-01649345.
- Christian Wolf, Eric Lombardi, Julien Mille, Oya Celiktutan, Mingyuan Jiu, Emre Dogan, Gonen Eren, Moez Baccouche, Emmanuel Dellandréa, Charles-Edmond Bichot et al. (2014). « Evaluation of video activity localizations integrating quality and quantity measurements ». Computer Vision and Image Understanding, vol. 127, pp. 14-30. doi : 10.1016/j.cviu.2014.06.014. HAL : hal-01283866

Thèse

Soutenu le 07/07/2017

Établissement :
INSA Lyon

Contact :
Atilla.Baskurt@iris.cnrs.fr

Conférences internationales avec comité de lecture

- Emre Dogan, Gonen Eren, Christian Wolf & Atilla Baskurt (2015). « Activity Recognition with Volume Motion Templates and Histograms of 3D Gradients ». International Conference on Image Processing, 30 septembre 2015, Quebec City (Canada). HAL : hal-01147957.

Autres

- Emre Dogan, Gonen Eren, Christian Wolf, Eric Lombardi & Atilla Baskurt (2017). « Multi-view pose estimation with mixtures-of-parts and adaptive viewpoint selection ». ArXiv : 1709.08527. HAL : hal-01593547.

Traitement des graphes massifs : partitionnement et requêtage agrégatif

Jury

Raphaël COUTURIER, PR, Université de Franche Comté
Karine ZEITOUNI, PR, Université de Versailles
Mohand BOUGHANEM, PR, Université Paul Sabatier
Angela BONIFATI, PR, Université Lyon 1
Hamamache KHEDDOUCI, PR, Université Lyon 1

Rapporteur
Rapporteuse
Examineur
Examinatrice
Directeur de thèse

Résumé

Avec l'avènement du "big data", de nombreuses répercussions ont eu lieu dans tous les domaines de la technologie de l'information, préconisant des solutions innovantes remportant le meilleur compromis entre coûts et précision. En théorie des graphes, où les graphes constituent un support de modélisation puissant qui permet de formaliser des problèmes allant des plus simples aux plus complexes, la recherche pour des problèmes NP-complet ou NP-difficile se tourne plutôt vers des solutions approchées, mettant ainsi en avant les algorithmes d'approximation et les heuristiques alors que les solutions exactes deviennent extrêmement coûteuses et impossible d'utilisation. Nous abordons dans cette thèse deux problématiques principales: dans un premier temps, le problème du partitionnement des graphes est abordé d'une perspective "big data", où les graphes massifs sont partitionnés en streaming. En effet, le partitionnement de graphe est une tâche cruciale et importante lors de la mise en place des graphes de données distribuées, pour la principale raison est qu'un bon partitionnement permet d'accélérer les calculs sur ces graphes de données distribuées. Dans ce contexte, nous étudions et proposons plusieurs modèles et heuristiques de partitionnement en streaming et nous évaluons leurs performances autant sur le plan théorique qu'empirique. Dans un second temps, nous nous intéressons au requêtage des graphes distribués ou partitionnés. En effet, une grande majorité de données est modélisée sous formes d'entités étiquetées et interconnectées, ce qui fait du requêtage des graphes distribués une tâche importante dans une large liste d'applications. Dans ce cadre, nous étudions la problématique de la "recherche agrégative dans les graphes" qui a pour but de répondre à des requêtes interrogeant plusieurs fragments de graphes et qui se charge de la reconstruction de la réponse finale tel que l'on obtient un "matching approché" avec la requête initiale.

Abstract

The advent of "big data" has tremendous repercussions on a broad range of data related domains, and it has impelled the use of novel techniques that achieve the best tradeoff between computational cost and precision. In the graph theory field, graphs represent a powerful tool for data and problem modeling where all kinds of problems, starting from the very simple to the very complicated, can be effectively formalized. To resolve NP-complete or NP-hard problems in this field, research is being directed to approximation algorithms and heuristic solutions rather than exact solutions, which exhibit a very high computational cost that makes them impossible to use for massive graph datasets. In this thesis, we tackle two main problems: first, the graph partitioning problem is studied in the context of "big data", where the focus is put on large graph streaming partitioning. In fact, the graph partitioning is an important and challenging problem when performing computation tasks over large distributed graphs, the reason is that a good partitioning leads to faster computations. We studied and proposed several streaming partitioning models and heuristics, and we studied their performances theoretically and

Ghizlane Echbarthi

Thèse

Soutenue le 23/10/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Hamamache.Kheddouci@liris.cnrs.fr

experimentally as well. The second problem that we tackle in this thesis, is the querying of partitioned/distributed graphs. In fact, querying graph datasets proves to be an important task as most of the real-life datasets are represented by networks of labeled entities. In this context, we study the problem of “aggregated graph search” that aims to answer queries on several graph fragments, and has the task to build a coherent final answer such that it forms an “approximate matching” to the initial query. We proposed a new method for “aggregated graph search”, and we studied its performances theoretically and experimentally on different real word graph datasets.

■ Publications

Conférences internationales avec comité de lecture

- Ghizlane Echbarthi & Hamamache Kheddouci (2014). « Partial Restreaming Approach for Massive Graph Partitioning. ». SITIS The 10th International Conference on SIGNAL IMAGE TECHNOLOGY & INTERNET BASED SYSTEMS, Marrakech (Maroc). HAL : hal-01282078.
- Ghizlane Echbarthi & Hamamache Kheddouci (2014). « Fractional greedy and partial restreaming partitioning: New methods for massive graph partitioning. ». Big Data Conference, Washington DC (États-Unis). HAL : hal-01282071.

Apprentissage de structure de modèles graphiques probabilistes : application à la classification multi-label

Maxime Gasse

Jury

Christophe GONZALES, PR, Université Paris 6
Jose M. PEÑA, Associate Professor, Linköping University
Elisa FROMONT, MC, Université Jean Monnet
Willem WAEGEMAN, PR, Ghent University
Véronique DELCROIX, MC, Université de Valenciennes
Céline ROBARDET, PR, INSA Lyon
Alexandre AUSSEM, PR, Université Lyon 1
Haytham ELGHAZEL, MC, Polytech Lyon

Rapporteur
Rapporteur
Examinatrice
Examineur
Examinatrice
Examinatrice
Directeur de thèse
Co-directeur de thèse

Résumé

Dans cette thèse, nous nous intéressons au problème spécifique de l'apprentissage de structure de modèles graphiques probabilistes, c'est-à-dire trouver la structure la plus efficace pour représenter une distribution, à partir seulement d'un ensemble d'échantillons $D = \{p(v)\}$. Dans une première partie, nous passons en revue les principaux modèles graphiques probabilistes de la littérature, des plus classiques (modèles dirigés, non-dirigés) aux plus avancés (modèles mixtes, cycliques etc.). Puis nous étudions particulièrement le problème d'apprentissage de structure de modèles dirigés (réseaux Bayésiens), et proposons une nouvelle méthode hybride pour l'apprentissage de structure, H2PC (Hybrid Hybrid Parents and Children), mêlant une méthode à base de contraintes (tests statistiques d'indépendance) et une méthode à base de score (probabilité postérieure de la structure).

Dans un second temps, nous étudions le problème de la classification multi-label, visant à prédire un ensemble de catégories (vecteur binaire y) pour un objet donné (vecteur x). Dans ce contexte, l'utilisation de modèles graphiques probabilistes pour représenter la distribution conditionnelle des catégories prend tout son sens, particulièrement dans le but minimiser une fonction coût complexe. Nous passons en revue les principales approches utilisant un modèle graphique probabiliste pour la classification multi-label (Probabilistic Classifier Chain, Conditional Dependency Network, Bayesian Network Classifier, Conditional Random Field, Sum-Product Network), puis nous proposons une approche générique visant à identifier une factorisation de $p(y|x)$ en distributions marginales disjointes, en s'inspirant des méthodes d'apprentissage de structure à base de contraintes. Nous démontrons plusieurs résultats théoriques, notamment l'unicité d'une décomposition minimale, ainsi que trois procédures quadratiques sous diverses hypothèses. Enfin, nous mettons en pratique ces résultats afin d'améliorer la classification multi-label avec les fonctions coût F-loss et zero-one loss.

Publications

Reuves internationales avec comité de lecture

- Maxime Gasse, Alex Aussem & Haytham Elghazel (2014). « A hybrid algorithm for Bayesian network structure learning with application to multi-label learning ». *Expert Systems with Applications*, vol. 41, n°15, pp. 6755-6772. doi : 10.1016/j.eswa.2014.04.032. ArXiv : 1506.05692. HAL : hal-01153255.
- Maxime Gasse, Alex Aussem & Haytham Elghazel (2012). « An Experimental Comparison of Hybrid Algorithms for Bayesian Network Structure Learning ». *Lecture notes in computer science*, vol. 7523, pp. 58-73. doi : 10.1007/978-3-642-33460-3_9. ArXiv : 1505.05004. HAL : hal-01122771.

Thèse

Soutenue le 13/01/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Alexandre.Aussem@liris.cnrs.fr

Conférences internationales avec comité de lecture

- Maxime Gasse & Alex Aussem (2016). « Identifying the irreducible disjoint factors of a multivariate probability distribution. ». Probabilistic Graphical Models, 9 septembre 2016, Lugano (Suisse), pp. 183-194. HAL : hal-01425447.
- Maxime Gasse & Alex Aussem (2016). « F-Measure Maximization in Multi-Label Classification with Conditionally Independent Label Subsets ». European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 23 septembre 2016, Riva del garda (Italie), pp. 619-631. doi : 10.1007/978-3-319-46128-1_39. HAL : hal-01425528.
- Maxime Gasse, Alex Aussem & Haytham Elghazel (2015). « On the Optimality of Multi-Label Classification under Subset Zero-One Loss for Distributions Satisfying the Composition Property ». International Conference on Machine Learning, 11 juillet 2015, Lille (France), pp. 2531-2539. HAL : hal-01234346.
- Alex Aussem, Pascal Caillet, Sarah Klemm, Maxime Gasse, Anne-Marie Schott & Michel Ducher (2014). « Analysis of risk factors of hip fracture with causal Bayesian networks ». International Work-Conference on Bioinformatics and Biomedical Engineering, 7 avril 2014, Grenade (Espagne), pp. 1074-1085. HAL : hal-01325484.
- Alex Aussem, Pascal Caillet, Sarah Klemm, Maxime Gasse, Anne-Marie Schott & Michel Ducher (2014). « Analysis of risk factors of hip fracture with causal Bayesian networks ». International Work-Conference on Bioinformatics and Biomedical Engineering, 7 avril 2014, Grenade (Espagne), pp. 1074-1085. HAL : hal-01326485.

Apprentissage multi-label ensembliste dans le contexte supervisé et semi-supervisé

Jury

Pascale KUNTZ, PR, Université de Nantes
Mustapha LEBBAH, MC, Université Paris 13
Elisa FROMONT, PR, Université Rennes 1
Alexandre AUSSEM, PR, Université Lyon 1
Haytham ELGHAZEL, MC, Université Lyon 1

Rapporteuse
Rapporteur
Examinatrice
Directeur de thèse
Co-Directeur de thèse

Ouadie Gharroudi

Résumé

L'apprentissage multi-label est un problème d'apprentissage supervisé où chaque instance peut être associée à plusieurs labels cibles simultanément. Il est omniprésent dans l'apprentissage automatique et apparaît naturellement dans de nombreuses applications du monde réel telles que la classification de documents, l'étiquetage automatique de musique et l'annotation d'images. Nous discutons d'abord pourquoi les algorithmes multi-label de l'état de l'art utilisant un comité de modèle souffrent de certains inconvénients pratiques. Nous proposons ensuite une nouvelle stratégie pour construire et agréger les modèles ensemblistes multi-label basés sur k-labels. Nous analysons ensuite en profondeur l'effet de l'étape d'agrégation au sein des approches ensemblistes multi-label et étudions comment cette agrégation influence les performances prédictive du modèle en fonction de la nature de fonction coût à optimiser. Nous abordons ensuite le problème spécifique de la sélection de variables dans le contexte multi-label en se basant sur le paradigme ensembliste. Trois méthodes de sélection de caractéristiques multi-label basées sur le paradigme des forêts aléatoires sont proposées. Ces méthodes diffèrent dans la façon dont elles considèrent la dépendance entre les labels dans le processus de sélection des variables. Enfin, nous étendons les problèmes de classification et de sélection de variables au cadre d'apprentissage semi-supervisé. Nous proposons une nouvelle approche de sélection de variables multi-label semi-supervisée basée sur le paradigme de l'ensemble. Le modèle proposé associe des principes issues de la co-training en conjonction avec une métrique interne d'évaluation d'importance des variables basée sur les out-of-bag. Testés de manière satisfaisante sur plusieurs données de référence, les approches développées dans cette thèse sont prometteuses pour une variété d'applications dans l'apprentissage multi-label supervisé et semi-supervisé.

Abstract

Multi-label learning is a specific supervised learning problem where each instance can be associated with multiple target labels simultaneously. Multi-label learning is ubiquitous in machine learning and arises naturally in many real-world applications such as document classification, automatic music tagging and image annotation. In this thesis, we formulate the multi-label learning as an ensemble learning problem in order to provide satisfactory solutions for both the multi-label classification and the feature selection tasks, while being consistent with respect to any type of objective loss function. We first discuss why the state-of-the-art single multi-label algorithms using an effective committee of multi-label models suffer from certain practical drawbacks. We then propose a novel strategy to build and aggregate k-label sets based committee in the context of ensemble multi-label classification. We then analyze the effect of the aggregation step within ensemble multi-label approaches in depth and investigate how this aggregation impacts the prediction performances with respect to the objective multi-label loss metric. We then address the specific problem of identifying relevant subsets

Thèse

Soutenue le 21/12/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Alexandre.Aussem@iris.cnrs.fr

of features - among potentially irrelevant and redundant features - in the multi-label context based on the ensemble paradigm. Three wrapper multi-label feature selection methods based on the Random Forest paradigm are proposed. These methods differ in the way they consider label dependence within the feature selection process. Finally, we extend the multi-label classification and feature selection problems to the semi-supervised setting and consider the situation where only few labelled instances are available. We propose a new semi-supervised multi-label feature selection approach based on the ensemble paradigm. The proposed model combines ideas from co-training and multi-label k- labelset committee construction in tandem with an inner out-of-bag label feature importance evaluation. Satisfactorily tested on several benchmark data, the approaches developed in this thesis show promise for a variety of applications in supervised and semi-supervised multi-label learning.

■ Publications

Conférences internationales avec comité de lecture

- O. Gharroudi, H. Elghazel, A. Aussem. « A Comparison of Multi-Label Feature Selection Methods Using the Random Forest Paradigm ». Canadian Conference on Artificial Intelligence, AI, Montréal. pp. 95-106. Springer. ISBN 978-3-319-06482-6. 2014.

Traitement joint de nuages de points et d'images pour l'analyse et la visualisation de formes 3D

Jury

Franck HÉTROUY-WHEELER, PR, Université de Strasbourg
Pascal MONASSE, CR, Ecole des Ponts ParisTech
Enric MEINHARDT-LLOPIS, MC, ENS Cachan
Luce MORIN, PR, INSA Rennes
Raphaëlle CHAINE, PR, Université Lyon 1
Julie DIGNE, CR, CNRS, Université Lyon 1
Pascal LEFEBVRE-ALBARET, Technodigit

Rapporteur
Rapporteur
Examineur
Examinatrice
Directrice de thèse
Co-directrice
Invité

Résumé

Au cours de la dernière décennie, les technologies permettant la numérisation d'espaces urbains ont connu un développement rapide. Des campagnes d'acquisition de données couvrant des villes entières ont été menées en utilisant des scanners LiDAR (Light Detection And Ranging) installés sur des véhicules mobiles. Les résultats de ces campagnes d'acquisition laser, représentant les bâtiments numérisés, sont des nuages de millions de points pouvant également contenir un ensemble de photographies. On s'intéresse ici à l'amélioration du nuage de points à l'aide des données présentes dans ces photographies. Cette thèse apporte plusieurs contributions notables à cette amélioration.

La position et l'orientation des images acquises sont généralement connues à l'aide de dispositifs embarqués avec le scanner LiDAR, même si ces informations de positionnement sont parfois imprécises. Pour obtenir un recalage précis d'une image sur un nuage de points, nous proposons un algorithme en deux étapes, faisant appel à l'information mutuelle normalisée et aux histogrammes de gradients orientés. Cette méthode permet d'obtenir une pose précise même lorsque les estimations initiales sont très éloignées de la position et de l'orientation réelles.

Une fois ces images recalées, il est possible de les utiliser pour inférer la couleur de chaque point du nuage en prenant en compte la variabilité des points de vue. Pour cela, nous nous appuyons sur la minimisation d'une énergie prenant en compte les différentes couleurs associables à un point et les couleurs présentes dans le voisinage spatial du point.

Bien entendu, les différences d'illumination lors de l'acquisition des données peuvent altérer la couleur à attribuer à un point. Notamment, cette couleur peut dépendre de la présence d'ombres portées amenées à changer avec la position du soleil. Il est donc nécessaire de détecter et de corriger ces dernières. Nous proposons une nouvelle méthode qui s'appuie sur l'analyse conjointe des variations de la réflectance mesurée par le LiDAR et de la colorimétrie des points du nuage. En détectant suffisamment d'interfaces ombre/lumière nous pouvons caractériser la luminosité de la scène et la corriger pour obtenir des scènes sans ombre portée.

Le dernier problème abordé par cette thèse est celui de la densification du nuage de points. En effet la densité locale du nuage de points est variable et parfois insuffisante dans certaines zones. Nous proposons une approche applicable directement par la mise en œuvre d'un filtre bilatéral joint permettant de densifier le nuage de points en utilisant les données des images.

Maximilien Guislain

Thèse

Soutenue le 19/10/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Raphaelle.Chaine@liris.cnrs.fr

Abstract

Recent years saw a rapid development of city digitization technologies. Acquisition campaigns covering entire cities are now performed using LiDAR (Light Detection And Ranging) scanners embedded aboard mobile vehicles. These acquisition campaigns yield point clouds, composed of millions of points, representing the buildings and the streets, and may also contain a set of images of the scene. The subject developed here is the improvement of the point cloud using the information contained in the camera images. This thesis introduces several contributions to this joint improvement.

The position and orientation of acquired images are usually estimated using devices embedded with the LiDAR scanner, even if this information is inaccurate. To obtain the precise registration of an image on a point cloud, we propose a two-step algorithm which uses both Mutual Information and Histograms of Oriented Gradients. The proposed method yields an accurate camera pose, even when the initial estimations are far from the real position and orientation. Once the images have been correctly registered, it is possible to use them to color each point of the cloud while using the variability of the point of view. This is done by minimizing an energy taking into account the different colors associated with a particular point and the potential colors of its neighbors. Illumination changes can also impact the color assigned to a point. Notably, this color can be affected by cast shadows. These cast shadows are changing with the sun position, it is therefore necessary to detect and correct them. We propose a new method that analyzes the joint variation of the reflectance value obtained by the LiDAR and the color of the points. By detecting enough interfaces between shadow and light, we are able to characterize the luminance of the scene and to remove the cast shadows. The last point developed in this thesis is the densification of a point cloud. Indeed, the local density of a point cloud varies and is sometimes insufficient in certain areas. We propose a directly applicable approach, using a joint bilateral filter, to increase the density of a point cloud using multiple images.

Publications

Revues internationales avec comité de lecture

- Maximilien Guislain, Julie Digne, Raphaëlle Chaine & Gilles Monnier (2017). « Fine scale image registration in large-scale urban LIDAR point sets ». *Computer Vision and Image Understanding*, vol. 157, pp. 90-102. doi : 10.1016/j.cviu.2016.12.004. HAL : hal-01468091.

Conférences internationales avec comité de lecture

- Maximilien Guislain, Julie Digne, Raphaëlle Chaine, Dimitri Kudelski & Pascal Lefebvre-Albaret (2016). « Detecting and Correcting Shadows in Urban Point Clouds and Image Collections ». 2016 International Conference on 3D Vision (3DV), Oct 2016, Stanford, United States, 28 octobre 2016, Stanford (États-Unis). HAL : hal-01393998.

Autres

- Maximilien Guislain, Julie Digne, Raphaëlle Chaine & Gilles Monnier (2016). « Recalage d'image dans des nuages de points de scènes urbaines ». Actes des journées du Groupe de Travail en Modélisation Géométrique (GTMG) 2016. HAL : hal-01320263.

Analyse automatisée de la circulation automobile par caméras de surveillance routière

Jury

Philippe JOLY, Université Paul Sabatier Toulouse 3
Lucile ROSSI, Université de Corse
Fatima BOUMGHAR-OULEBSIR, Univ. Houari Boumédiène
Chalermchai PANYADEE, Université de Maejo
Serge MIGUET, Université Lumière Lyon 2
Mihaela SCUTURICI, Université Lumière Lyon 2

Rapporteur
Rapporteuse
Examinatrice
Examineur
Directeur de thèse
Co-directrice de thèse

Résumé

Cette thèse s'inscrit dans le contexte de l'analyse vidéo du trafic routier. Dans certaines grandes villes, des centaines de caméras produisent de très grandes quantités de données, impossible à manipuler sans traitement automatique. Notre principal objectif est d'aider les opérateurs humains en analysant automatiquement les données vidéo. Pour aider les contrôleurs de la circulation à prendre leurs décisions, il est important de connaître en temps réel, l'état du trafic (nombre de véhicules et vitesse des véhicules sur chaque segment de voie), mais aussi de disposer de statistiques temporelles tout au long de la journée, de la semaine, de la saison ou de l'année. Les caméras ont été déployées depuis longtemps pour le trafic et pour d'autres fins de surveillance, car elles fournissent une source d'information riche pour la compréhension humaine. L'analyse vidéo peut désormais apporter une valeur ajoutée aux caméras en extrayant automatiquement des informations pertinentes. De cette façon, la vision par ordinateur et l'analyse vidéo deviennent de plus en plus importantes pour les systèmes de transport intelligents (intelligent transport systems : ITSs). L'une des problématiques abordées dans cette thèse est liée au comptage automatique de véhicules. Pour être utile, un système de surveillance vidéo doit être entièrement automatique et capable de fournir, en temps réel, l'information qui concerne le comportement de l'objet dans la scène. Nous pouvons obtenir ces renseignements sur la détection et le suivi des objets en mouvement dans les vidéos, ce qui a été un domaine largement étudié. Néanmoins, la plupart des systèmes d'analyse automatique par vidéo ont des difficultés à gérer les situations particulières.

Aujourd'hui, il existe de nombreux défis à résoudre tels que les occultations entre les différents objets, les arrêts longs, les changements de luminosité, etc. qui conduisent à des trajectoires incomplètes. Dans la chaîne de traitements que nous proposons, nous nous sommes concentrés sur l'extraction automatique de statistiques globales dans les scènes de vidéosurveillance routière. Notre chaîne de traitements est constituée par les étapes suivantes : premièrement, nous avons évalué différentes techniques de segmentation de vidéos et de détection d'objets en mouvement. Nous avons choisi une méthode de segmentation basée sur une version paramétrique du mélange de gaussiennes appliquée sur une hiérarchie de blocs, méthode qui est considérée actuellement comme l'un des meilleurs procédés pour la détection d'objets en mouvement. Nous avons proposé une nouvelle méthodologie pour choisir les valeurs optimales des paramètres d'un algorithme permettant d'améliorer la segmentation d'objets en utilisant des opérations morphologiques. Nous nous sommes intéressés aux différents critères permettant d'évaluer la qualité d'une segmentation, résultant d'un compromis entre une bonne détection des objets en mouvement, et un faible nombre de fausses détections, par exemple causées par des changements d'illumination, des reflets ou des bruits d'acquisition. Deuxièmement, nous effectuons une classification des objets, basée sur les descripteurs de Fourier,

Kannikar Intawong

Thèse

Soutenue le 27/09/2017

Établissement :

Université Lumière Lyon 2

Contact :

Serge.Miguet@liris.cnrs.fr

et nous utilisons ces descripteurs pour éliminer les objets de type piétons ou autres et ne conserver que les véhicules. Troisièmement, nous utilisons un modèle de mouvement et un descripteur basé sur les couleurs dominantes pour effectuer le suivi des objets extraits. En raison des difficultés mentionnées ci-dessus, nous obtenons des trajectoires incomplètes, qui donneraient une information de comptage erronée si elles étaient exploitées directement. Nous proposons donc d'agréger les données partielles des trajectoires incomplètes et de construire une information globale sur la circulation des véhicules dans la scène. Notre approche permet la détection des points d'entrée et de sortie dans les séquences d'images. Nous avons testé nos algorithmes sur des données privées provenant du centre de contrôle du trafic de la ville de Chiang Mai en Thaïlande, ainsi que sur des données publiques du MIT (MIT, 2009). Sur ce dernier jeu de données, nous avons comparé les performances de nos algorithmes à des articles précédemment publiés et utilisant les mêmes données. Dans plusieurs situations, nous illustrons les améliorations apportées par notre méthode en termes de localisation de zones d'entrée / de sortie, et en termes de comptage des véhicules.

Mot clés : vidéo-surveillance, détection d'objets en mouvement, détection des zones entrée-sortie, mesures de qualité

Publications

Conférences internationales avec comité de lecture

- K. Intawong, M. Scuturici, S. Miguet. "A new pixel-based quality measure for segmentation algorithms integrating precision, recall and specificity". International Conference on Computer Analysis of Images and Patterns (CAIP), York. pp. 188-195. 2013.

Addressing Selfishness in the Design of Cooperative Systems

Jury

Harald KOSCH, Universität Passau, Germany	Rapporteur
Vivien QUEMA, Grenoble INP/ENSIMAG, France	Rapporteur
Julia LAWALL, Sorbonne Universités, UPMC, LIP6, France	Rapporteur
Sara BOUCHENAK, INSA Lyon, France	Examineur
Paola BONIZZONI, Università degli Studi di Milano-Bicocca	Examineur
Mariagrazia FUGINI, Politecnico di Milano, Italy	Examineur
Ernesto DAMIANI, Università degli Studi di Milano	Directeur de thèse
Sonia BEN MOKHTAR, INSA Lyon	Co-encadrante
Gianini GABRIELE, Università degli Studi di Milano	Co-encadrant

Abstract

Cooperative systems are at the basis of several mainstream Internet applications (ex.: peer-to-peer file-sharing and media streaming). Essential to their success is the voluntary contribution of resources (ex.: bandwidth, CPU, storage) from the participating nodes. However, as nodes are autonomous entities, they may be tempted to behave in a selfish manner by not contributing their fair share, potentially causing system performance degradation and instability. Addressing selfish nodes is, therefore, key to building efficient and reliable cooperative systems. Yet, it is a challenging task, as current techniques for analysing selfishness and designing effective countermeasures remain manual and time-consuming, requiring multi-domain expertise.

In this thesis, we aim to provide practical and conceptual tools to help system designers in dealing with selfish nodes. First, based on a comprehensive survey of existing work on selfishness, we develop a classification framework to identify and understand the most important selfish behaviours to focus on when designing a cooperative system.

Second, we propose RACOON, a unifying framework for the selfishness-aware design and configuration of cooperative systems. RACOON provides a semi-automatic methodology to integrate a given system with practical and finely tuned mechanisms to meet specified resilience and performance objectives, using game theory and simulations to predict the behaviour of the system when subjected to selfish nodes. An extension of the framework (RACOON++) is also proposed to improve the accuracy, flexibility, and usability of RACOON.

Finally, we propose SEINE, a framework for fast modelling and evaluation of various types of selfish behaviour in a given cooperative system. SEINE relies on a domain-specific language for describing the selfishness scenario to evaluate, and provides semi-automatic support for its implementation and study in a state-of-the-art simulator.

Publications

Conférences internationales avec comité de lecture

- Guido Lena Cota, Sonia Ben Mokhtar, Julia Lawall, Gilles Muller, Gabriele Gianini, Ernesto Damiani & Lionel Brunie (2015). « A Framework for the Design Configuration of Accountable Selfish-Resilient Peer-to-Peer Systems ». SRDS 2015 - 34th International Symposium on Reliable Distributed Systems, 1 octobre 2015, Montreal (Canada). HAL : hal-01250717.

Guido Lena Cota

Thèse

Soutenue le 24/03/2017

Établissement :
INSA Lyon & UNIVERSITÀ DEGLI
STUDI DI MILANO

Contact :
Lionel.Brunie@iris.cnrs.fr



Conception et développement d'interactions immersives pour jeux sérieux

Jury

Philippe FUCHS, PR, Mines ParisTech
Sébastien GEORGE, PR
Jean-Pierre JESSEL, PR
Domitile LOURDEAUX, MC
Daniel MESTRE, DR
Audrey SERNA, MC

Examineur
Directeur de thèse
Rapporteur
Rapporteuse
Examineur
Co-Encadrante

Résumé

Cette thèse s'inscrit dans le domaine de l'ingénierie des Environnements Virtuels pour l'Apprentissage Humain (EVAH) et a été financée par le projet ANR JEN.lab. Les jeux épistémiques numériques (JEN) sont une catégorie des jeux sérieux dédiés à la résolution de problèmes complexes, pluridisciplinaires et non-déterministes. L'objectif de ces JEN est de proposer des situations d'apprentissage authentiques du point de vue des interactions qu'elles permettent, les apprenants pouvant alors construire et ancrer des connaissances dans leur contexte d'usage.

La première partie de la thèse est consacrée à une méthodologie permettant de concevoir des JEN aux interactions enrichies sur le plan spatial, temporel et social, nommés jeux épistémiques numériques pervasifs (JENP). Des expérimentations ont été menées en milieu écologique et ont montré un apport pédagogique par une augmentation de l'engagement des apprenants. La conception des prototypes expérimentés a mis en évidence que le développement informatique des interactions immersives d'un EVAH demeure actuellement réservé à des spécialistes en Réalité Mixte ou Virtuelle (RMV). Ainsi la seconde partie de la thèse a été dédiée à la proposition d'un environnement de développement nommé MIREEDGE pouvant être utilisé par des développeurs non spécialistes en interactions immersives. Une interface de programmation visuelle permet aux développeurs de modéliser des interactions sans connaissance spécifique en RMV par réutilisation de blocs. Les principaux blocs sont destinés à représenter visuellement une fonctionnalité ou un algorithme d'un périphérique de RMV. Des blocs logiques permettent aux développeurs d'établir un algorithme d'une interaction particulière par composition. Enfin d'autres blocs permettent de relier le nouvel algorithme à d'autres méthodes existantes dans le projet où l'interaction sera intégrée. Des expérimentations ont permis d'évaluer l'efficacité et l'efficacité de MIREEDGE pour le profil des développeurs RMV ainsi que celui des développeurs non-initiés à la RMV. Les deux profils de développeurs sont complémentaires dans MIREEDGE, les premiers proposant de nouveaux blocs d'interactions à l'ensemble de la communauté, les seconds, en plus grand nombre, pouvant mener à un assemblage optimal. Les développeurs RMV ayant une nécessité à poursuivre leur algorithme au-delà de la programmation visuelle, MIREEDGE permet une génération automatique de code fournissant le script correspondant et totalement rééritable dans leur moteur de jeu. L'approche MIREEDGE, dans son ensemble, s'inscrit dans une architecture logicielle permettant le partage rapide d'interactions selon un méta-modèle. Son indépendance logicielle permet de garantir une large compatibilité avec les environnements de développement et les périphériques, ouvrant ainsi des perspectives sur l'étude de logiques d'interactions universelles.

Mots-clés : Jeux sérieux, Jeux épistémiques numériques, Interactions immersives, Développeur de jeux sérieux, Programmation visuelle, Moteur de jeu, Développeur en réalité virtuelle.

Guillaume Loup

Thèse

Soutenue le 04/12/2017

Établissement :
INSA Lyon

Contact :
Audey.Serna@iris.cnrs.fr



Transfert de connaissances pour la classification des images

Jury

Nicolas THOME, PR, Conservatoire national des arts et métiers Rapporteur
Krystian MIKOLAJCZYK, PR, Imperial College London Rapporteur
Nicole VINCENT, PR, Université Paris Descartes Examinatrice
Amaury HABRARD, PR, Université Jean Monnet Invité
Liming CHEN, PR, Ecole Centrale de Lyon Directeur de thèse
Alexandre SAIDI, MC, Ecole Centrale de Lyon Co-directeur de thèse

Résumé

Lors de l'apprentissage d'un modèle de classification pour un nouveau domaine cible avec seulement une petite quantité d'échantillons de formation, l'application des algorithmes d'apprentissage automatiques conduit généralement à des classifieurs surdimensionnés avec de mauvaises compétences de généralisation. D'autre part, recueillir un nombre suffisant d'échantillons de formation étiquetés manuellement peut s'avérer très coûteux. Les méthodes de transfert d'apprentissage visent à résoudre ce type de problèmes en transférant des connaissances provenant d'un domaine source associé qui contient beaucoup plus de données pour faciliter la classification dans le domaine cible. Selon les différentes hypothèses sur le domaine cible et le domaine source, l'apprentissage par transfert peut être classé en trois catégories: apprentissage par transfert inductif, apprentissage par transfert transducteur (adaptation du domaine) et apprentissage par transfert non surveillé. Nous nous concentrons sur le premier qui suppose que la tâche cible et la tâche source sont différentes mais liées. Plus précisément, nous supposons que la tâche cible et la tâche source sont des tâches de classification, tandis que les catégories cible et les catégories source sont différentes mais liées. Nous proposons deux méthodes différentes pour aborder ce problème.

Dans le premier travail, nous proposons une nouvelle méthode d'apprentissage par transfert discriminatif, à savoir DTL (Discriminative Transfer Learning), combinant une série d'hypothèses faites à la fois par le modèle appris avec les échantillons de cible et les modèles supplémentaires appris avec des échantillons des catégories sources. Plus précisément, nous utilisons le résidu de reconstruction creuse comme discriminant de base et améliorons son pouvoir discriminatif en comparant deux résidus d'un dictionnaire positif et d'un dictionnaire négatif. Sur cette base, nous utilisons des similitudes et des dissemblances en choisissant des catégories sources positivement corrélées et négativement corrélées pour former des dictionnaires supplémentaires. Une nouvelle fonction de coût basée sur la statistique de Wilcoxon-Mann-Whitney est proposée pour choisir les dictionnaires supplémentaires avec des données non équilibrées. En outre, deux processus de Boosting parallèles sont appliqués à la fois aux distributions de données positives et négatives pour améliorer encore les performances du classificateur. Sur deux bases de données de classification d'images différentes, la DTL proposée surpasse de manière constante les autres méthodes de l'état de l'art du transfert de connaissances, tout en maintenant un temps d'exécution très efficace.

Dans le deuxième travail, nous combinons le pouvoir du transport optimal (OT) et des réseaux de neurones profonds (DNN) pour résoudre le problème ITL. Plus précisément, nous proposons une nouvelle méthode pour affiner conjointement un réseau de neurones avec des données source et des données cibles. En ajoutant une fonction de perte du transfert optimal (OT loss) entre les prédictions du classificateur source et cible comme une

Ying Lu

Thèse

Soutenue le 09/11/2017

Établissement :

Ecole Centrale de Lyon

Contact :

Liming.Chen@liris.cnrs.fr

contrainte sur le classificateur source, le réseau JTLN (Joint Transfer Learning Network) proposé peut effectivement apprendre des connaissances utiles pour la classification cible à partir des données source. En outre, en utilisant différents métriques comme matrice de coût pour la fonction de perte du transfert optimal, JTLN peut intégrer différentes connaissances antérieures sur la relation entre les catégories cibles et les catégories sources. Nous avons effectué des expérimentations avec JTLN basées sur Alexnet sur les jeux de données de classification d'image et les résultats vérifient l'efficacité du JTLN proposé. A notre connaissances, ce JTLN proposé est le premier travail à aborder ITL avec des réseaux de neurones profond (DNN) tout en intégrant des connaissances antérieures sur la relation entre les catégories cible et source.

Abstract

When learning a classification model for a new target domain with only a small amount of training samples, brute force application of machine learning algorithms generally leads to over-fitted classifiers with poor generalization skills. On the other hand, collecting a sufficient number of manually labeled training samples may prove very expensive. Transfer Learning methods aim to solve this kind of problems by transferring knowledge from related source domain which has much more data to help classification in the target domain. Depending on different assumptions about target domain and source domain, transfer learning can be further categorized into three categories: Inductive Transfer Learning, Transductive Transfer Learning (Domain Adaptation) and Unsupervised Transfer Learning. We focus on the first one which assumes that the target task and source task are different but related. More specifically, we assume that both target task and source task are classification tasks, while the target categories and source categories are different but related. We propose two different methods to approach this ITL problem.

In the first work we propose a new discriminative transfer learning method, namely DTL, combining a series of hypotheses made by both the model learned with target training samples, and the additional models learned with source category samples. Specifically, we use the sparse reconstruction residual as a basic discriminant, and enhance its discriminative power by comparing two residuals from a positive and a negative dictionary. On this basis, we make use of similarities and dissimilarities by choosing both positively correlated and negatively correlated source categories to form additional dictionaries. A new Wilcoxon-Mann-Whitney statistic based cost function is proposed to choose the additional dictionaries with unbalanced training data. Also, two parallel boosting processes are applied to both the positive and negative data distributions to further improve classifier performance. On two different image classification databases, the proposed DTL consistently outperforms other state-of-the-art transfer learning methods, while at the same time maintaining very efficient runtime.

In the second work we combine the power of Optimal Transport and Deep Neural Networks to tackle the ITL problem. Specifically, we propose a novel method to jointly fine-tune a Deep Neural Network with source data and target data. By adding an Optimal Transport loss (OT loss) between source and target classifier predictions as a constraint on the source classifier, the proposed Joint Transfer Learning Network (JTLN) can effectively learn useful knowledge for target classification from source data. Furthermore, by using different kind of metric as cost matrix for the OT loss, JTLN can incorporate different prior knowledge about the relatedness between target categories and source categories.

We carried out experiments with JTLN based on Alexnet on image classification datasets and the results verify the effectiveness of the proposed JTLN in comparison with standard consecutive fine-tuning. To the best of our knowledge, the proposed JTLN is the first work to tackle ITL with Deep Neural Networks while incorporating prior knowledge on relatedness between target and source categories. This Joint Transfer Learning with OT loss is general and can also be applied to other kind of Neural Networks.

Publications

Conférences internationales avec comité de lecture

- Learning Visual Categories through a Sparse Representation Classifier based Cross-Category Knowledge Transfer. Y. Lu, L. Chen, S.AS Saidi, Z. Zhang, Y. Wang. IEEE International Conference on Image Processing 2014, Paris. pp. 165-169. 2014.

Autres conférences

- Learning Visual Categories through a Sparse Representation Classifier based Cross-Category Knowledge Transfer. Y. Lu, L. Chen, S.AS Saidi, Z. Zhang, Y. Wang. IEEE International Conference on Image Processing 2014, Paris. pp. 1-5. 2014.



Une approche orientée service pour la recherche sémantique de contenus multimédias

Jury

Latifa BABA HAMED, PR, université d'Oran	Examinatrice
Mohand BOUGHANEM, PR, université de Toulouse	Rapporteur
Sidi M. BENSLIMANE, PR, école sup. en Informatique de Sidi Bel Abbès	Rapporteur
Mohammed Amine CHIKH, PR, université de Tlemcen	Examineur
Youssef AMGHAR, PR, INSA Lyon	Directeur de thèse
Azeddine CHIKH, PR, université de Tlemcen	Directeur de thèse

Résumé

Les sources de données multimédias provenant de divers domaines (médical, tourisme, commerce, art et culture, etc.) sont devenues incontournables sur le web. L'accès à ces sources multimédias dans les systèmes distribués pose de nouveaux problèmes en raison de nombreux paramètres : volumétrie, diversité des interfaces, format de représentation, localisation, etc. En outre, l'exigence de plus en plus forte des utilisateurs et des applications à vouloir intégrer la sémantique dans la recherche d'information pose de nouvelles questions à résoudre. Pour prendre en compte cette nouvelle complexité, nous nous intéressons dans notre travail de recherche aux solutions d'intégration de données basées sur les services web.

Dans cette thèse, nous proposons une approche orientée service pour la recherche sémantique de contenus multimédia. Nous avons appelé cette approche SeSaM (Semantic Search of Multimedia content). SeSaM repose sur la définition d'un nouveau type de services accédant aux contenus multimédias, qui est les services MaaS (Multimedia as a Services). Elle est basée sur un processus en deux phases : description et découverte des services MaaS. En ce qui concerne la description de services MaaS, nous avons défini le langage SA4MaaS (Semantic Annotation for MaaS services), qui est une extension de SAWSDL (recommandation W3C). L'idée principale de ce langage est l'intégration, en plus de la sémantique métier, de la sémantique de l'information multimédia dans la description des services MaaS.

En ce qui concerne la découverte de services MaaS, nous avons proposé un nouveau matchmaker MaaS-MX (MaaS services Matchmaker) adapté au modèle de description des MaaS. MaaS-MX est composé de deux étapes primordiales : appariement métier et appariement multimédia. L'appariement métier consiste à comparer la description métier des services et de la requête, tandis que l'appariement multimédia compare la description multimédia des services et de la requête.

L'approche a été prototypée et évaluée dans deux domaines différents : médical et tourisme. Les résultats indiquent que l'utilisation de l'appariement métier et l'appariement multimédia a considérablement amélioré les performances des systèmes de recherche de données multimédias.

Mots clés : Services web sémantiques, Recherche d'information, Données multimédias, Description de services, Découverte de services.

Abstract

Multimedia data sources from various fields (medical, tourism, trade, art and culture, etc.) became essential on the web. Accessing to multimedia data in distributed systems poses new challenges due to many system

Sid Ahmed Djalal
Midouni

Thèse

Soutenue le 08/07/2017

Établissement :
université Abou Bekr Blekaid
(cotutelle INSA Lyon)

Contact :
Youssef.Amghar@liris.cnrs.fr

parameters: volume, diversity of interfaces, representation format, location, etc. In addition, the growing needs of users and applications to incorporate semantics in the information retrieval pose new issues. To take into account this new complexity, we are interested in our research of data integration solutions based on web services.

In this thesis, we propose an approach-oriented service for the semantic search of multimedia content. We called this approach SeSaM (Semantic Search of Multimedia content). SeSaM is based on the definition of a new pattern of services to access multimedia content, which is the MaaS services (Multimedia as a Services). It is based on a two-phase process: description and discovery of MaaS services. As for the MaaS services description, we have defined the SA4MaaS language (Semantic Annotation for MaaS services), which is an extension of SAWSDL (W3C recommendation). The main idea of this language is the integration, in addition to business domain semantic, of multimedia information semantics in the MaaS services description.

As for the MaaS service discovery, we have proposed a new matchmaker MaaS-MX (MaaS services Matchmaker) adapted to the MaaS services description model. MaaS-MX is composed of two essential steps: domain matching and multimedia matching. Domain matching consists in comparing the business domain description of MaaS services and the query, whereas multimedia matching compares the multimedia description of MaaS services and the query.

The approach has been implemented and evaluated in two different domains: medical and tourism. The results indicate that using both domain and multimedia matching considerably improves the performance of multimedia data retrieving systems.

Keywords: Semantic web services, Information retrieval, Multimedia Data, Web service description, Web service discovery.

Publications

Revue internationale avec comité de lecture

- Sid Ahmed Djallal Midouni, Youssef Amghar & Azeddine Chikh (2016). « A semantic annotation for MaaS services: the SA4MaaS approach ». International Journal of Metadata, Semantics and Ontologies, vol. 11, n°2, pp. 82-92. HAL : hal-01435367.
- Sid Ahmed Djallal Midouni, Amghar Youssef & Chikh Azeddine (2016). « Multimedia Data Retrieving based on SOA Architecture ». Journal of Web Engineering, Vol.15 No.3&4, pp. 339-360. HAL : hal-01302391.

Conférences internationales avec comité de lecture

- Sid Ahmed Djallal Midouni, Amghar Youssef & Chikh Azeddine (2014). « A full service approach for multimedia content retrieval ». Model and Data Engineering, 24 septembre 2014, Larnaca (Chypre). doi : 10.1007/978-3-319-11587-0. HAL : hal-01213915.

Utilisation de méthodes de décomposition pour les problèmes du plus grand sous-graphe commun et de la somme coloration

Jury

Simon DE GIVRY, CR, centre de recherches de Toulouse
Chu Min LI, PR, université de Picardie Jules Verne, Amiens
Christophe LECOUTRE, PR, université d'Artois
Christine SOLNON, PR, INSA Lyon, Villeurbanne
Samba NDOJH NDIAYE, MC, université Lyon 1

Rapporteur
Rapporteur
Examineur
Directrice de thèse
Co-directeur

Résumé

L'objectif de cette thèse est, d'un point de vue général, de concevoir et d'évaluer des méthodes de décomposition applicables à des problèmes d'optimisation sous contraintes. Deux problèmes d'optimisation, en particulier, ont été considérés: le problème du plus grand sous-graphe commun, qui consiste à trouver la plus grande partie commune entre deux graphes, et le problème de la somme coloration, dans lequel un graphe doit être colorié d'une façon minimisant une somme de poids.

Le problème du plus grand sous-graphe commun (MCIS) est connu pour être très difficile. Il trouve des applications notamment en biologie, en chimie et en traitement d'image, où il survient lorsqu'on compare des objets structurés afin de mesurer leur taux de similarité. L'emploi d'une méthode de décomposition et de techniques de parallélisation est fortement recommandée compte tenu de la difficulté de ce problème. Cependant, les techniques de décomposition existantes s'avèrent peu adaptées au problème du MCIS : certaines génèrent des sous-problèmes très déséquilibrés, tandis que d'autres, comme la décomposition arborescente, sont tout bonnement inapplicables à ce problème.

Afin de permettre la décomposition de tels problèmes, Chmeiss et al. ont proposé une approche, la TR-decomposition, qui agit à un plus bas niveau : la microstructure du problème. Cette approche n'avait encore jamais été employée pour le problème du MCIS. Nous l'évaluons dans ce contexte, en cherchant à réduire la taille globale de l'espace de recherche et en nous intéressant aux possibilités de parallélisation. Nous introduisons également une étape de postdécomposition visant à limiter les redondances entre sous-problèmes.

Le second problème étudié est celui de la somme coloration. Il, s'agit d'une variante NP-difficile du très connu problème de coloration de graphes. Comme dans la plupart des problèmes de coloration, il est nécessaire d'affecter à chaque sommet du graphe considéré une couleur, tout en veillant à ce qu'aucuns sommets voisins n'utilisent la même couleur. Ce qui distingue le problème de la somme coloration est que chaque couleur possède un poids. Le but est de minimiser la somme des poids associés aux couleurs utilisées par chaque sommet du graphe. Cela produit des instances généralement plus difficiles que le problème de coloration classique, qui, lui, implique simplement une minimisation du nombre total de couleurs utilisées.

Peu de méthodes complètes ont été proposées pour ce problème. On trouve cependant dans la littérature, notamment, un modèle de programmation par contraintes (PPC), une approche de type branch and bound, ainsi qu'un modèle de programmation linéaire en nombres entiers (PLNE).

Nous avons étudié de manière poussée les capacités de la PPC à résoudre le problème de somme coloration, tout en cherchant des moyens d'augmenter

Maël Minot

Thèse

Soutenue le 19/12/2017

Établissement :
INSA Lyon

Contact :
Christine.Solnon@liris.cnrs.fr

ses performances. Nous avons également étudié une combinaison de la PLNE et de la PPC visant à combiner les points forts de ces deux approches grandement complémentaires. Nous nous sommes inspirés de la très classique technique du « backtracking » borné par une décomposition arborescente (BTD). Nous employons une décomposition arborescente, mais celle-ci a une hauteur bornée à 1. La PPC est utilisée pour énumérer les affectations cohérentes de la racine, et, pour chacune de ces affectations, un sous-problème est trivialement obtenu pour chaque feuille de la décomposition. La PLNE est utilisée pour résoudre à l'optimum chacun de ces sous-problèmes, chacun d'entre eux étant parfaitement indépendant.

Afin de tirer parti de la complémentarité de nos approches, nous avons développée une approche portfolio. Le solveur qui en résulte est capable de choisir, pour chaque instance considérée, une des approches disponibles automatiquement en s'appuyant sur un certain nombre de traits observés dans l'instance.

Abstract

The objective of this thesis is, from a general standpoint, to design and evaluate decomposition methods for solving constrained optimisation problems. Two optimisation problems in particular are considered: the maximum common induced subgraph problem, in which the largest common part between two graphs is to be found, and the sum colouring problem, where a graph must be coloured in a way that minimises a sum of weights induced by the employed colours.

The maximum common subgraph (MCIS) problem is notably difficult, with a strong applicability in domains such as biology, chemistry and image processing, where the need to measure the similarity between structured objects represented by graphs may arise. The outstanding difficulty of this problem makes it strongly advisable to employ a decomposition method, possibly coupled with a parallelisation of the solution process. However, existing decomposition methods are not well suited to solve the MCIS problem: some lead to a poor balance between subproblems, while others, like tree decomposition, are downright inapplicable.

To enable the structural decomposition of such problems, Chmeiss et al. proposed an approach, TR-decomposition, acting at a low level: the microstructure of the problem. This approach had yet to be applied to the MCIS problem. We evaluate it in this context, aiming at reducing the size of the search space while also enabling parallelisation. Moreover, we introduce a post-decomposition step that focuses on alleviating redundancies between subproblems.

The second problem that caught our interest is the sum colouring problem. It is an NP-hard variant of the widely known classical graph colouring problem. As in most colouring problems, it basically consists in assigning colours to the vertices of a given graph while making sure no neighbour vertices use the same colour. In the sum colouring problem, however, each colour is associated with a weight. The objective is to minimise the sum of the weights of the colours used by every vertex. This leads to generally harder instances than the classical colouring problem, which simply requires to use as few colours as possible.

Only a few exact methods have been proposed for this problem. Among them stand notably a constraint programming (CP) model, a branch and bound approach, as well as an integer linear programming (ILP) model.

We led an in-depth investigation of CP's capabilities to solve the sum colouring problem, while also looking into ways to make it more efficient. Additionally, we evaluated a combination of integer linear programming and constraint programming, with the intention of conciliating the strong points of these highly complementary approaches. We took inspiration from the classical backtracking bounded by tree decomposition (BTD) approach. We employ a tree decomposition with a strictly bounded height. Constraint programming is used to enumerate consistent assignment of the root cluster. For each of these assignments, one subproblem is trivially obtained for each leaf cluster. Subproblems are then independently solved to optimality using ILP.

We then derive profit from the complementarity of our approaches by developing a portfolio approach. The resulting solver is able to choose one of the considered approaches automatically by relying on a number of features extracted from each instance.

Publications

Conférences internationales avec comité de lecture

- Maël Minot, Samba Ndojh Ndiaye & Christine Solnon (2017). « Combining CP and ILP in a tree decomposition of bounded height for the sum colouring problem ». CPAIOR 2017, 8 juin 2017, Padova (Italie). HAL : hal-01447818.
- Maël Minot, Samba Ndojh Ndiaye & Christine Solnon (2016). « Using CP and ILP with tree decomposition to solve the sum colouring problem ». Doctoral program of CP 2016, 9 septembre 2016, Toulouse (France). HAL : hal-01366291.
- Maël Minot, Samba Ndojh Ndiaye & Christine Solnon (2015). « A Comparison of Decomposition Methods for the Maximum Common Subgraph Problem ». 27th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), 11 novembre 2015, Vierti sul Mare (Italie). HAL : hal-01191605.
- Maël Minot & Samba Ndojh Ndiaye (2014). « Searching for a maximum common induced subgraph by decomposing the compatibility graph ». Bridging the Gap Between Theory and Practice in Constraint Solvers, CP2014-Workshop, 8 septembre 2014, Lyon (France), pp. 1-17. HAL : hal-01301095.

Conférences nationales avec comité de lecture

- Maël Minot, Samba Ndojh Ndiaye & Christine Solnon (2016). « Évaluation d'approches complètes pour le problème de somme coloration ». Douzièmes Journées Francophones de Programmation par Contraintes (JFPC 2016), 17 juin 2016, Montpellier (France). HAL : hal-01309350.
- Maël Minot, Samba Ndojh Ndiaye & Christine Solnon (2015). « Recherche d'un plus grand sous-graphe commun par décomposition du graphe de compatibilité ». Onzièmes Journées Francophones de Programmation par Contraintes (JFPC), 24 juin 2015, Bordeaux (France). HAL : hal-01163322.



Isomorphisme de sous-graphes dans les graphes de données massifs

Jury

Lynda TAMINE-LECHANI, PR, Université Paul Sabatier
Ludovic LIETARD, MC, Université Rennes 1
Alexandre TERMIER, PR, Université Rennes 1
Hamida SEBA, MC, Université Lyon 1

Rapporteuse
Rapporteur
Examineur
Directrice de thèse

Résumé

L'interrogation de graphes de données est un problème fondamental qui connaît un grand intérêt, en particulier pour les données structurées massives où les graphes constituent une alternative prometteuse aux bases de données relationnelles pour la modélisation des grandes masses de données. Cependant, l'interrogation des graphes de données est différente et plus complexe que l'interrogation des données relationnelles à base de tables. La tâche principale impliquée dans l'interrogation de graphes de données est la recherche d'isomorphisme de sous-graphes qui est un problème NP-complet.

La recherche d'isomorphisme de sous-graphes est un problème très important impliqué dans divers domaines comme la reconnaissance de formes, l'analyse des réseaux sociaux, la biologie, etc. Il consiste à énumérer les sous-graphes d'un graphe de données qui correspondent à un graphe requête. Les solutions les plus connues de ce problème sont basées sur le retour arrière (backtracking). Elles explorent un grand espace de recherche, ce qui entraîne un coût de traitement élevé, notamment dans le cas de données massives.

Pour réduire le temps et la complexité en espace mémoire dans la recherche d'isomorphisme de sous-graphes, nous proposons d'utiliser des graphes compressés. Dans notre approche, la recherche d'isomorphisme de sous-graphes est réalisée sur une représentation compressée des graphes sans les décompresser. La compression des graphes s'effectue en regroupant les sommets en super-sommets. Ce concept est connu dans la théorie des graphes par la décomposition modulaire. Il sert à générer une représentation en arbre d'un graphe qui met en évidence des groupes de sommets qui ont les mêmes voisins. Avec cette compression, nous obtenons une réduction substantielle de l'espace de recherche et par conséquent, une économie significative dans le temps de traitement.

Nous proposons également une nouvelle représentation des sommets du graphe, qui simplifie le filtrage de l'espace de recherche. Ce nouveau mécanisme appelé "compact neighborhood Index (CNI)" encode l'information de voisinage autour d'un sommet en un seul entier. Cet encodage du voisinage réduit la complexité du temps de filtrage de cubique à quadratique. Ce qui est considérable pour les données massifs.

Nous proposons également un algorithme de filtrage itératif qui repose sur les caractéristiques des CNIs pour assurer un élagage global de l'espace de recherche.

Nous avons évalué nos approches sur plusieurs datasets et nous les avons comparés avec les algorithmes de l'état de l'art.

Chemseddine Nabti

Thèse

Soutenue le 15/12/2017

Établissement :
Université Claude Bernard Lyon1

Contact :
Hamida.Seba@liris.cnrs.fr

Abstract

Querying graph data is a fundamental problem that witnesses an increasing interest especially for massive structured data where graphs come as a promising alternative to relational databases for big data modeling. However, querying graph data is different and more complex than querying relational table-based data. The main task involved in querying graph data is subgraph isomorphism search which is an NP-complete problem. Subgraph isomorphism search, is an important problem which is involved in various domains such as pattern recognition, social network analysis, biology, etc. It consists to enumerate the subgraphs of a data graph that match a query graph. The most known solutions of this problem are backtracking-based. They explore a large search space which results in a high computational cost when we deal with massive graph data.

To reduce time and memory space complexity of subgraph isomorphism search. We propose to use compressed graphs. In our approach, subgraph isomorphism search is achieved on compressed representations of graphs without decompressing them. Graph compression is performed by grouping vertices into super vertices. This concept is known, in graph theory, as modular decomposition. It is used to generate a tree representation of a graph that highlights groups of vertices that have the same neighbors. With this compression we obtain a substantial reduction of the search space and consequently a significant saving in the processing time.

We also propose a novel encoding of vertices that simplifies the filtering of the search space. This new mechanism is called compact neighborhood Index (CNI).

A CNI distills all the information around a vertex in a single integer. This simple neighborhood encoding reduces the time complexity of vertex filtering from cubic to quadratic which is considerable for big graphs. We propose also an iterative local global filtering algorithm that relies on the characteristics of CNIs to ensure a global pruning of the search space.

We evaluated our approaches on several real-world datasets and compared them with the state of the art algorithms.

Publications

Revue internationale avec comité de lecture

- Chemseddine Nabti & Hamida Seba (2017). « Querying massive graph data: A compress and search approach ». *Future Generation Computer Systems*, vol. 74, pp. 63-75. doi : 10.1016/j.future.2017.04.005. HAL : hal-01546073.

Conférences internationales avec comité de lecture

- Chemseddine Nabti & Hamida Seba (2017). « Querying massive graph data: A compress and search approach ». *International Conference on Internet of Things, Big Data and Security - IoTBDS 2017*, 26 avril 2017, Porto (Portugal). HAL : hal-01546080.
- Chemseddine Nabti & Hamida Seba (2016). « Subgraph Isomorphism Search in Massive Graph Databases ». *The International Conference on Internet of Things and Big Data – IoTBD 2016*, 25 avril 2016, Rome (Italie). HAL : hal-01313922.

Jury

Anne RUAS, IR (ICPEF), IFSTTAR
Romain RAFFIN, MC, Aix-Marseille Université
Marjorie MUSY, DR, Cerema Ouest
William PUECH, PR, Université de Montpellier
Benoit GOURGAND, PR, DINSI, Grand Lyon Métropole
Gilles GESQUIERE, PR, Université Lumière Lyon 2

Rapportrice
Rapporteur
Examinatrice
Examineur
Invité
Directeur de thèse

Résumé

Les modélisations 3D de ville se multiplient à travers le monde et deviennent aujourd'hui accessibles grâce à la volonté des communes de les proposer librement. Il est ainsi aujourd'hui possible d'accéder à plusieurs milliers de kilomètres carrés de territoires urbains modélisés en 3D.

Nous présentons dans cette thèse un ensemble de méthodes permettant d'enrichir un modèle virtuel 3D de ville, de l'organiser afin de faciliter son utilisation, puis de l'analyser en détectant les changements entre deux millésimes ou en mesurant son ambiance visuelle selon les besoins de l'utilisateur. Nous proposons dans un premier temps un algorithme permettant de compléter un modèle 3D en y adjoignant une définition sémantique grâce au cadastre, ainsi qu'une méthode de calcul de la canopée végétale 3D par croisement de données. Puis nous proposons des stratégies d'organisation du modèle 3D de la ville selon des critères géométriques et sémantiques afin de faciliter son parcours. Nous présentons ensuite des méthodes comparant deux millésimes de données représentant un même territoire dans le but de produire un unique modèle temporel contenant un ensemble de versions. Enfin, nous cherchons à mesurer l'ambiance visuelle de la ville via l'analyse de la composition d'un champ de vision ainsi que de l'ensoleillement et des ombres portées présentes sur un territoire. Ces mesures peuvent être liées à des données externes afin de proposer de multiples interprétations en fonction des besoins de l'utilisateur.

Tous ces travaux se font dans un contexte d'interopérabilité et de genericité puisqu'ils ont pour objectif de pouvoir être utilisés avec des jeux de données provenant du monde entier. Nous basons donc nos méthodes sur l'utilisation de standards internationaux aussi bien pour les données en entrée que pour les résultats en sortie. Il est ainsi possible, dans le cadre d'une approche voulue dans un contexte de pluralité scientifique au sein du LabEx Intelligences des Mondes Urbains (IMU), de mettre à disposition de la communauté les résultats mais aussi nos algorithmes proposés au sein d'un logiciel développé en open source.

Abstract

3D virtual models of cities are multiplying throughout the world and now become accessible thanks to the will of the communes to propose them freely. Today, it is possible to access several thousand square kilometres of urban territories modelled in 3D.

We present in this thesis a set of methods to enrich a 3D virtual city model, organize it to facilitate its use and then analyse it by detecting changes between two vintages or by measuring its visual atmosphere according to the user's needs. We first propose an algorithm to complete a 3D model by adding a semantic definition thanks to the cadastre, and a method to compute the

Frédéric Pedrinis

Thèse

Soutenue le 17/10/2017

Établissement :

Université Lumière Lyon2

Contact :

Gilles.Gesquiere@liris.cnrs.fr

3D plant canopy by crossing data. Then we propose strategies to organize the 3D model of the city according to geometric and semantic criteria in order to facilitate its browsing. We then present methods comparing two vintages of data representing the same territory in order to produce a single temporal model containing a set of versions. Finally, we try to measure the visual atmosphere of the city by analysing the composition of a field of vision as well as the amount of sunshine and the shadows on a territory. These measurements can be linked to external data in order to propose multiple interpretations according to the user's needs.

All this work is done in a context of interoperability and genericity since it aims to be used with datasets from all over the world. We therefore base our methods on the use of international standards for both input and output data. It is thus possible, in a context of scientific plurality within the LabEx IMU, to make available to the community the results but also our algorithms proposed within an open source developed software.

Publications

Revues internationales avec comité de lecture

- Frédéric Pedrinis & Gilles Gesquière (2017). « Reconstructing 3D Building Models with the 2D Cadastre for Semantic Enhancement ». *Lecture Notes in Geoinformation and Cartography*, pp. 119-135. doi : 10.1007/978-3-319-25691-7_7. HAL : hal-01383234.
- Frédéric Pedrinis, Maxime Morel & Gilles Gesquière (2015). « Change Detection of Cities ». *Lecture Notes in Geoinformation and Cartography*. doi : 10.1007/978-3-319-12181-9_8. HAL : hal-01214108.

Conférences internationales avec comité de lecture

- Vincent Jaillot, Frédéric Pedrinis, Sylvie Servigne & Gilles Gesquière (2017). « A generic approach for sunlight and shadow impact computation on large city models ». *25th International Conference on Computer Graphics, Visualization and Computer Vision 2017*, 2 juin 2017, Pilsen (République Tchèque), 10 pages. HAL : hal-01559175.
- Florence Jacquinod, Frédéric Pedrinis, Jérémy Edert & Gilles Gesquière (2016). « Automated Production of Interactive 3D Temporal Geovisualizations so as to Enhance Flood Risk Awareness ». *UDMV 2016*, 8 décembre 2016, Liège (Belgique). HAL : hal-01413338.
- Jérémy Gaillard, Alexandre Vienne, Rémi Baume, Frédéric Pedrinis, Adrien Peytavie & Gilles Gesquière (2015). « Urban data visualisation in a web browser ». *Web3D 2015*, 21 juin 2015, Heraklion (Grèce), pp. 81-88. doi : 10.1145/2775292.2775302. HAL : hal-01196834.

Subspace clustering on static datasets and dynamic data streams using bio-inspired algorithms

Sergio Peignier

Jury

Wolfgang BANZHAF, CR, Michigan State University	Rapporteur
Stefan KRAMER, CR, Johannes Gutenberg University	Rapporteur
Sylvie GALICHET, PR, Polytech Annecy-Chambéry	Examinatrice
Ruggero PENSA, Assistant Professor, Univ. degli Studi di Torino	Examineur
Susan STEPNEY, CR, University of York	Examinatrice
Sébastien VEREL, MC, Université du Littoral Côte d'Opale	Examineur
Christophe RIGOTTI, MC, INSA-Lyon	Directeur de thèse
Guillaume BESLON, PR, INSA-Lyon	Directeur de thèse

Résumé

Les récents progrès techniques ont facilité l'acquisition massive de données décrites par un grand nombre de propriétés mesurables (jeu de données à forte dimensionalité). De plus, le développement de nouvelles technologies a permis l'acquisition continue des données temporelles, fournissant aux utilisateurs des flux de données potentiellement infinis. Dans ces deux cas, les algorithmes traditionnels de clustering s'avèrent souvent insuffisants. En effet, les mesures de similarité, couramment utilisées par les techniques de clustering traditionnelles, rencontrent des limites lorsqu'elles sont utilisées dans des espaces à forte dimensionalité. Ce phénomène conduit à une dégradation de la qualité du modèle de clustering obtenu. D'autre part, les grands volumes des flux de données ne permettent pas d'utiliser des techniques qui nécessitent l'exécution de plusieurs passes sur le jeu de données. Pour surmonter ces problèmes, une variété de nouvelles approches ont été proposées dans la littérature. Une tâche importante qui a été étudiée dans le contexte de données à forte dimensionalité est la tâche connue sous le nom de subspace clustering. Le subspace clustering est généralement reconnu comme étant plus compliqué que le clustering standard, étant donné que cette tâche vise à détecter des groupes d'objets similaires entre eux (clusters), et en même temps elle vise à trouver les sous-espaces où apparaissent ces similitudes. De plus, le subspace clustering, ainsi que le clustering traditionnel ont été récemment étendus au traitement de flux de données en mettant à jour les modèles de clustering de façon incrémentale. Les différents algorithmes qui ont été proposés dans la littérature, reposent sur des bases algorithmiques très différentes. Parmi ces approches, les algorithmes évolutifs ont été sous-explorés, même si ces techniques se sont avérées très utiles pour traiter d'autres problèmes NP-difficiles. L'objectif de cette thèse a été de tirer parti des nouvelles connaissances issues de l'évolution afin de concevoir des algorithmes évolutifs qui traitent le problème du subspace clustering sur des jeux de données statiques ainsi que sur des flux de données dynamiques. Chameleoclust, le premier algorithme développé au cours de ce projet, tire partie du grand degré de liberté fourni par des éléments bio-inspirés tels qu'un génome de longueur variable, l'existence d'éléments fonctionnels et non fonctionnels et des opérateurs de mutation incluant des réarrangements chromosomiques. KymeroClust, le deuxième algorithme conçu dans cette thèse, est un algorithme de k-medians qui repose sur un mécanisme évolutif de grande importance: la duplication et la divergence des gènes. SubMorphoStream, le dernier algorithme développé ici, aborde le problème du subspace clustering sur des flux de données dynamiques. Cet algorithme repose sur deux mécanismes qui jouent un rôle important dans l'adaptation rapide des bactéries à des environnements changeants: à savoir, l'amplification de gènes et l'absorption de matériel génétique externe. Ces algorithmes ont été comparés aux principales techniques de l'état de l'art, et ont obtenu des résultats compétitifs. En outre, deux applications appelées

Thèse

Soutenue le 27/07/2017

Établissement :
INSA Lyon

Contact :
Guillaume.Beslon@liris.cnrs.fr

EvoWave et EvoMove ont été développés pour évaluer la capacité de ces algorithmes à résoudre des problèmes réels. EvoWave est une application qui analyse de signaux Wi-Fi pour détecter des contextes différents. EvoMove est un compagnon musical artificiel qui produit des sons basés sur le clustering des mouvements d'un danseur, capturés à l'aide des capteurs de mouvement.

Abstract

Recent technical advances have facilitated the massive acquisition of data described by a large number of measurable properties (high dimensional datasets). New technologies have also enabled the continuous acquisition of data over time, providing users with possibly infinite data streams. The analysis of both high dimensional and streaming data by means of traditional clustering algorithm turn out to be troublesome. In the context of high dimensional data, common similarity measures used by clustering techniques tend to be less meaningful, leading to a degradation of the clustering quality. Moreover, in the case of data streams, the large volume of data does not allow to run several passes on the dataset. In order to overcome these problems, a variety of new approaches has been proposed in the literature. An important task that has been investigated in the context of high dimensional data is subspace clustering. This data mining task is recognized as more general and complicated than standard clustering, since it aims to detect groups of similar objects called clusters, and at the same time to find the subspaces where these similarities appear. Furthermore, subspace clustering approaches as well as traditional clustering ones have recently been extended to deal with data streams by updating clustering models in an incremental way. The different algorithms that have been proposed in the literature, rely on very different algorithmic foundations. Among these approaches, evolutionary algorithms have been under-explored, even if these techniques have proven to be valuable addressing other NP-hard problems. The aim of this thesis was to take advantage of new knowledge from evolutionary biology in order to conceive evolutionary subspace clustering algorithms for static datasets and dynamic data streams. Chameleoclust, the first algorithm developed in this work, takes advantage of the large degree of freedom provided by bio-like features such as a variable genome length, the existence of functional and non-functional elements and mutation operators including chromosomal rearrangements. KymeroClust, our second algorithm, is a k-medians based approach that relies on the duplication and the divergence of genes, a cornerstone evolutionary mechanism. SubMorphoStream, the last one, tackles the subspace clustering task over dynamic data streams. It relies on two important mechanisms that favor fast adaptation of bacteria to changing environments, namely gene amplification and foreign genetic material uptake. All these algorithms were compared to the main state-of-the-art techniques, obtaining competitive results. Results suggest that these algorithms are useful complementary tools in the analyst toolbox. In addition, two applications called EvoWave and EvoMove have been developed to assess the capacity of these algorithms to address real world problems. EvoWave is an application that handles the analysis of Wi-Fi signals to detect different contexts. EvoMove, the second one, is a musical companion that produces sounds based on the clustering of dancer moves captured using motion sensors.

Publications

Revue internationale avec comité de lecture

- Jonas Abernot, Guillaume Beslon, Simon Hickinbotham, Sergio Peignier & Christophe Rigotti (2017). « Evolving Instrument Based on Symbiont-Host

Metaphor: a Commensal Computation ». *Journal of Creative Music Systems*, vol. 2, n°1, pp. 1-10. doi : 10.5920/JCMS.2017.11. HAL : hal-01596573.

Conférences internationales avec comité de lecture

- Sergio Peignier, Jonas Abernot, Christophe Rigotti & Guillaume Beslon (2017). « EvoMove: Evolutionary-based living musical companion ». *European Conference on Artificial Life (ECAL)*, 8 septembre 2017, Villeurbanne (France), pp. 1-8. HAL : hal-01569091.
- Jonas Abernot, Guillaume Beslon, Simon Hickinbotham, Sergio Peignier & Christophe Rigotti (2016). « A Commensal Architecture for Evolving Living Instruments ». *Conference on Computer Simulation of Musical Creativity*, 19 juin 2016, Huddersfield (Royaume-Uni), pp. 1-8. HAL : hal-01368034.
- Sergio Peignier, Christophe Rigotti & Guillaume Beslon (2015). « Subspace Clustering for all Seasons ». *EvoEvo Workshop (satellite workshop of ECAL 2015)*, 24 juillet 2015, York (Royaume-Uni), pp. 1-3. HAL : hal-01252793.
- Sergio Peignier, Christophe Rigotti & Guillaume Beslon (2015). « Subspace Clustering Using Evolvable Genome Structure ». *Genetic and Evolutionary Computation Conference (GECCO)*, 11 juillet 2015, Madrid (Espagne), pp. 575-582. HAL : hal-01199136.



Introducing Privacy in Current Web Search Engines

Jury

Claude CASTELLUCCIA, DR, INRIA Grenoble
Rüdiger KAPITZA, PR, TU Braunschweig
Harald KOSCH, PR, Universität Passau
Catherine BERRUT, PR, Université Grenoble Alpes
Mathias LUX, MC, Universität Klagenfurt
Lionel BRUNIE, PR, INSA Lyon
Michael GRANITZER, PR, Universität Passau
Sonia BEN MOKHTAR, CR, CNRS / LIRIS

Rapporteur
Rapporteur
Examineur
Examinatrice
Examineur
Directeur de thèse
Directeur de thèse
Co-directrice de thèse

Résumé

Au cours des dernières années les progrès technologiques permettant de collecter, stocker et traiter d'importantes quantités de données pour un faible coût, ont soulevés de sérieux problèmes concernant la vie privée. La protection de la vie privée concerne de nombreux domaines, en particulier les sites internet fréquemment utilisés comme les moteurs de recherche (ex. : Google, Bing, Yahoo!). Ces services permettent aux utilisateurs de retrouver efficacement du contenu sur Internet en exploitant leurs données personnelles. Dans ce contexte, développer des solutions pour permettre aux utilisateurs d'utiliser ces moteurs de recherche tout en protégeant leurs vies privées est devenu primordial.

Dans cette thèse, nous introduisons SimAttack, une attaque contre les solutions protégeant la vie privée de l'utilisateur dans ses interactions avec les moteurs de recherche. Cette attaque vise à retrouver les requêtes initialement envoyées par l'utilisateur. Nous avons montré avec cette attaque que trois mécanismes représentatifs de l'état de l'art ne sont pas satisfaisants pour protéger la vie privée des utilisateurs.

Par conséquent, nous avons développé PEAS, un nouveau mécanisme de protection qui améliore la protection de la vie privée de l'utilisateur. Cette solution repose sur deux types de protection : cacher l'identité de l'utilisateur (par une succession de deux serveurs) et masquer sa requête (en la combinant avec des fausses requêtes). Afin de générer des fausses requêtes réalistes, PEAS se base sur les précédentes requêtes envoyées par les utilisateurs du système.

Pour finir, nous présenterons des mécanismes permettant d'identifier la sensibilité des requêtes. Notre objectif est d'adapter les mécanismes de protection existants pour protéger uniquement les requêtes sensibles, et ainsi économiser des ressources (ex. : CPU, mémoire vive). Nous avons développé deux modules pour identifier les requêtes sensibles. En déployant ces modules sur des mécanismes de protection existants, nous avons établi qu'ils permettent d'améliorer considérablement leurs performances.

Abstract

During the last few years, the technological progress in collecting, storing and processing a large quantity of data for a reasonable cost has raised serious privacy issues. Privacy concerns many areas, but is especially important in frequently used services like search engines (e.g., Google, Bing, Yahoo!). These services allow users to retrieve relevant content on the Internet by exploiting their personal data. In this context, developing solutions to enable users to use these services in a privacy-preserving way is becoming increasingly important.

Albin Petit

Thèse

Soutenue le 15/03/2017

Établissement :
INSA Lyon

Contact :
Lionel.Brunie@liris.cnrs.fr

In this thesis, we introduce SimAttack an attack against existing protection mechanism to query search engines in a privacy-preserving way. This attack aims at retrieving the original user query. We show with this attack that three representative state-of-the-art solutions do not protect the user privacy in a satisfactory manner.

We therefore develop PEAS a new protection mechanism that better protects the user privacy. This solution leverages two types of protection: hiding the user identity (with a succession of two nodes) and masking users' queries (by combining them with several fake queries). To generate realistic fake queries, PEAS exploits previous queries sent by the users in the system.

Finally, we present mechanisms to identify sensitive queries. Our goal is to adapt existing protection mechanisms to protect sensitive queries only, and thus save user resources (e.g., CPU, RAM). We design two modules to identify sensitive queries. By deploying these modules on real protection mechanisms, we establish empirically that they dramatically improve the performance of the protection mechanisms.

Publications

Revue internationale avec comité de lecture

- Albin Petit, Thomas Cerqueus, Antoine Boutet, Sonia Ben Mokhtar, David Coquil, Lionel Brunie & Harald Kosch (2016). « SimAttack: private web search under fire ». *Journal of Internet Services and Applications*, p. 17. doi : 10.1186/s13174-016-0044-x. HAL : hal-01304320.

Conférences internationales avec comité de lecture

- Jérémie Decouchant, Sonia Ben Mokhtar, Albin Petit & Vivien Quéma (2016). « PAG: Private and Accountable Gossip ». *ICDCS - International Conference on Distributed Computing Systems*, 30 juin 2016, Nara (Japon). doi : 10.1109/ICDCS.2016.34. HAL : hal-01368911.
- Albin Petit, Thomas Cerqueus, Sonia Ben Mokhtar, Lionel Brunie & Harald Kosch (2015). « PEAS: Private, Efficient and Accurate Web Search ». *14th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 22 août 2015, Helsinki (Finlande). HAL : hal-01159179.
- Albin Petit, Sonia Ben Mokhtar, Lionel Brunie & Harald Kosch (2014). « Towards Efficient and Accurate Privacy Preserving Web Search ». *Proceedings of the 9th Workshop on Middleware for Next Generation Internet Computing*, 8 décembre 2014, Bordeaux, France (France), 1:1-1:6. doi : 10.1145/2676733.2676734. HAL : hal-01301099..

Single image super-resolution based on neural networks for text and face recognition

Jury

Thierry CHATEAU, PR, Univ. de Clermont-Auvergne
Jean-Philippe THIRAN, PR, EPFL
Christine GUILLEMOT, PR, INRIA
Christian VIARD-GAUDIN, PR, Univ. de Nantes
Christophe GARCIA, PR, INSA de Lyon
Moez BACCOUCHE, Ingénieur de recherche, Orange

Rapporteur
Rapporteur
Examinatrice
Examineur
Directeur de thèse
Co-encadrant

Résumé

Cette thèse porte sur les méthodes de super-résolution (SR) pour l'amélioration des performances des systèmes de reconnaissance automatique (OCR, reconnaissance faciale). Les méthodes de Super-Résolution (SR) permettent de générer des images haute résolution (HR) à partir d'images basse résolution (BR). Contrairement à un rééchantillonnage par interpolation, elles restituent les hautes fréquences spatiales et compensent les artéfacts (flou, crénelures). Parmi elles, les méthodes d'apprentissage automatique telles que les réseaux de neurones artificiels permettent d'apprendre et de modéliser la relation entre les images BR et HR à partir d'exemples. Ce travail démontre l'intérêt des méthodes de SR à base de réseaux de neurones pour les systèmes de reconnaissance automatique. Les réseaux de neurones à convolutions sont particulièrement adaptés puisqu'ils peuvent être entraînés à extraire des caractéristiques non-linéaires bidimensionnelles pertinentes tout en apprenant la correspondance entre les espaces BR et HR. Sur des images de type documents, la méthode proposée permet d'améliorer la précision en reconnaissance de caractère de +7.85 points par rapport à une simple interpolation. La création d'une base d'images annotée et l'organisation d'une compétition internationale (ICDAR2015) ont souligné l'intérêt et la pertinence de telles approches. Pour les images de visages, les caractéristiques faciales sont cruciales pour la reconnaissance automatique. Une méthode en deux étapes est proposée dans laquelle la qualité de l'image est d'abord globalement améliorée, pour ensuite se focaliser sur les caractéristiques essentielles grâce à des modèles spécifiques. Les performances d'un système de vérification faciale se trouvent améliorées de +6.91 à +8.15 points. Enfin, pour le traitement d'images BR en conditions réelles, l'utilisation de réseaux de neurones profonds permet d'absorber la variabilité des noyaux de flous caractérisant l'image BR, et produire des images HR ayant des statistiques naturelles sans connaissance du modèle d'observation exact.

Abstract

This thesis is focussed on super-resolution (SR) methods for improving automatic recognition system (Optical Character Recognition, face recognition) in realistic contexts. SR methods allow to generate high resolution images from low resolution ones. Unlike upsampling methods such as interpolation, they restore spatial high frequencies and compensate artefacts such as blur or jaggy edges. In particular, example-based approaches learn and model the relationship between low and high resolution spaces via pairs of low and high resolution images. Artificial Neural Networks are among the most efficient systems to address this problem. This work demonstrate the interest of SR methods based on neural networks for improved automatic recognition systems. By adapting the data, it is possible to train such Machine Learning algorithms to produce high-resolution images. Convolutional Neural Networks are especially efficient as they are trained to simultaneously extract relevant non-linear features while learning the mapping between low

Clément Peyrard

Thèse

Soutenu le 15/12/2017

Établissement :
INSA Lyon

Contact :
Christophe.garcia@liris.cnrs.fr

and high resolution spaces. On document text images, the proposed method improves OCR accuracy by +7.85 points compared with simple interpolation. The creation of an annotated image dataset and the organisation of an international competition (ICDAR2015) highlighted the interest and the relevance of such approaches. Moreover, if a priori knowledge is available, it can be used by a suitable network architecture. For facial images, face features are critical for automatic recognition. A two step method is proposed in which image resolution is first improved, followed by specialised models that focus on the essential features. An off-the-shelf face verification system has its performance improved from +6.91 up to +8.15 points. Finally, to address the variability of real-world low-resolution images, deep neural networks allow to absorb the diversity of the blurring kernels that characterise the low-resolution images. With a single model, high-resolution images are produced with natural image statistics, without any knowledge of the actual observation model of the low-resolution image.

Publications

Conférences internationales avec comité de lecture et actes

- A comparison between Multi-Layer Perceptrons and Convolutional Neural Networks for Text Image Super-Resolution . C. Peyrard, F. Mamalet, C. Garcia. Dans 10th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP 2015), Berlin. 2015.

Architecture de monitoring déclaratif durable : optimisation énergétique des interactions entre requêtes applicatives orientées service et réseau de capteurs sans fil – Application aux bâtiments intelligents

Ozgun Pinarer

Jury

Jean-Philippe BABAU, PR, Université de Brest
Atilla BASKURT, PR, INSA Lyon
Alain BOUJU, MC, Université de La Rochelle
Yann GRIPAY, MC, INSA Lyon
Thérèse LIBOUREL, PR Emérite, Université de Montpellier
François PINET, DR, IRSTEA Clermont-Ferrand

Rapporteur
Directeur de thèse
Examineur
Co-encadrant
Examinatrice
Rapporteur

Résumé

Des rapports de recherche et d'analyse soulignent qu'aujourd'hui, la consommation d'énergie élevée des bâtiments est un problème majeur dans les pays développés. Ces rapports montrent que les systèmes de gestion d'énergie du bâtiment et les réseaux de capteurs sans fil sont importants pour l'efficacité énergétique des bâtiment. Dans la littérature, les systèmes existants de gestion des bâtiments intelligents se concentrent sur la consommation d'énergie du bâtiment et du matériel déployé à l'intérieur / à l'extérieur, ainsi que sur les problématique de communication réseau. Ils adoptent des configurations statiques pour les capteurs sans fil, et les modèles proposés sont adaptés pour une application unique.

Dans cette thèse, nous proposons une architecture de monitoring déclarative durable qui se concentre sur l'optimisation énergétique des interactions entre les requêtes applicatives orientées services et le réseau de capteurs sans fil. Nous considérons le système de monitoring comme un ensemble d'applications qui exploitent les mesures de capteurs en temps réel. Ces applications peuvent être configurés dynamiquement par les utilisateurs. Dans notre approche, nous prenons un point de vue « données »: les applications sont exprimés de façon déclarative comme un ensemble de requêtes continues sur les flux de données des capteurs. Pour atteindre notre objectif d'optimisation énergétique de l'architecture de monitoring, nous formalisons la configuration du capteur et adaptons l'acquisition et de transmission de données selon les besoins réels des applications. Nous présentons une architecture de monitoring complète et un algorithme qui gère la configuration dynamique des capteurs. Nous introduisons une plateforme qui gère des capteurs sans fil physiques ou simulées.

Abstract

Recent researches and analysis reports highlight that high energy consumption of buildings is a major problem in developed countries. They show that building energy management systems and wireless sensor networks are important for energy efficiency of buildings. In the literature, existing smart building management systems focus on energy consumption of the building and of hardware deployed inside/outside of the building, and on network communication issues. They adopt static configurations for wireless sensor devices, and proposed models are fitted to a single application.

In this thesis, we propose a sustainable declarative monitoring architecture that focuses on the energy optimisation of interactions between application service oriented queries and wireless sensor devices. We consider the monitoring system as a set of applications that exploit sensor measures in real time. These applications can be configured dynamically by users. In our approach, we take a data point of view: applications are declaratively expressed as a set of continuous queries on sensor data streams. To achieve

Thèse

Soutenue le 15/12/2017

Établissement :
INSA Lyon

Contact :
Atilla.Baskurt@iris.cnrs.fr

our objective of energy aware optimization of the monitoring architecture, we formalize sensor device configuration and fit data acquisition and data transmission to actual applications requirements. We present a complete monitoring architecture and an algorithm that handles dynamic sensor configuration. We introduce a platform that manages physical and simulated wireless sensor devices.

Publications

Revues internationales avec comité de lecture

- Ozgun Pinarer, Yann Gripay, Sylvie Servigne, Ozgovde Atay & Atilla Baskurt (2017). « Dynamic energy-aware sensor configuration in multi-application monitoring systems ». *Pervasive and Mobile Computing*, vol. 41, pp. 192-204. doi : 10.1016/j.pmcj.2017.08.005. HAL : hal-01585263.

Conférences internationales avec comité de lecture

- Sylvie Servigne, Yann Gripay, Ozgun Pinarer, John Samuel, Atay Ozgovde & Jacques Jay (2016). « Heterogeneous Sensor Data Exploration and Sustainable Declarative Monitoring Architecture: Application to Smart Building ». *First International Conference on Smart Data and Smart Cities, 30th UDMS*, 9 septembre 2016, Split (Croatie), pp. 97-104. doi : 10.5194/isprs-annals-IV-4-W1-97-2016. HAL : hal-01386110.
- Ozgun Pinarer, Yann Gripay, Sylvie Servigne & Atay Ozgovde (2016). « Energy Enhancement of Multi-application Monitoring Systems for Smart Buildings ». *CAiSE'16: Conference on Advanced Information Systems Engineering - EnBIS: Energy-awareness and Big Data Management in Information Systems*, 17 juin 2016, Ljubljana (Slovénie), pp 131-142. doi : 10.1007/978-3-319-39564-7_14. HAL : hal-01334900.

Étude de l'évolution des micro-organismes bactériens par des approches de modélisation et de simulation informatique

Charles Rocabert

Jury

Samuel BERNARD, CR, CNRS
Guillaume BESLON, PR, INSA Lyon
Bahram HOUCHMANDZADEH, DR, CNRS
Carole KNIBBE, MC, INSA Lyon
Jean-Baptiste MOURET, CR, INRIA
Olivier TENAILLON, DR, INSERM
Karine VAN DONINCK, PR, University of Namur

Examineur
Directeur de thèse
Rapporteur
Directrice de thèse
Examineur
Rapporteur
Examinatrice

Résumé

Variation et sélection sont au cœur de l'évolution Darwinienne. Cependant, ces deux mécanismes dépendent de processus eux-mêmes façonnés par l'évolution. Chez les micro-organismes, qui font face à des environnements souvent variables, ces propriétés adaptatives sont particulièrement bien exploitées, comme le démontrent de nombreuses expériences en laboratoire. Chez ses organismes, l'évolution semble donc avoir optimisé sa propre capacité à évoluer, un processus que nous nommons évolution de l'évolution (EvoEvo). La notion d'évolution de l'évolution englobe de nombreux concepts théoriques, tels que la variabilité, l'évolvabilité, la robustesse ou encore la capacité de l'évolution à innover (open-endedness). Ces propriétés évolutives des micro-organismes, et plus généralement de tous les organismes vivants, sont soupçonnées d'agir à tous les niveaux d'organisation biologique, en interaction ou en conflit, avec des conséquences souvent complexes et contre-intuitives. Ainsi, comprendre l'évolution de l'évolution implique l'étude de la trajectoire évolutive de micro-organismes – réels ou virtuels –, et ce à différents niveaux d'organisation (génome, interactome, population,...).

L'objectif de ce travail de thèse a été de développer et d'étudier des modèles mathématiques et numériques afin de lever le voile sur certains aspects de l'évolution de l'évolution. Ce travail multidisciplinaire, car impliquant des collaborations avec des biologistes expérimentateur-riche-s, des bio-informaticien-ne-s et des mathématicien-ne-s, s'est divisé en deux parties distinctes, mais complémentaires par leurs approches : (i) l'extension d'un modèle historique en génétique des populations – le modèle géométrique de Fisher – afin d'étudier l'évolution du bruit phénotypique en sélection directionnelle, et (ii) le développement d'un modèle d'évolution *in silico* multi-échelles permettant une étude plus approfondie de l'évolution de l'évolution. Cette thèse a été financée par le projet européen EvoEvo (FP7-ICT-610427).

Abstract

Variation and selection are the two core processes of Darwinian Evolution. Yet, both are directly regulated by many processes that are themselves products of evolution. Microorganisms efficiently exploit this ability to dynamically adapt to new conditions. Thus, evolution seems to have optimized its own ability to evolve, as a primary means to react to environmental changes. We call this process evolution of evolution (EvoEvo). EvoEvo covers several aspects of evolution, encompassing major concepts such variability, evolvability, robustness, and open-endedness. Those phenomena are known to affect all levels of organization in bacterial populations. Indeed, understanding EvoEvo requires to study organisms experiencing evolution, and to decipher the evolutive interactions between all the components of the biological system of interest (genomes, biochemical networks, populations,...).

Thèse

Soutenu le 17/11/2017

Établissement :
INSA Lyon

Contact :
Guillaume.Beslon@liris.cnrs.fr

The objective of this thesis was to develop and exploit mathematical and numerical models to tackle different aspects of EvoEvo, in order to produce new knowledge on this topic, in collaboration with partners from diverse fields, including experimental biology, bioinformatics, mathematics and also theoretical and applied informatics. To this aim, we followed two complementary approaches: (i) a population genetics approach to study the evolution of phenotypic noise in directional selection, by extending Fisher's geometric model of adaptation, and (ii) a digital genetics approach to study multi-level evolution. This work was funded by the EvoEvo project, under the European Commission (FP7-ICT-610427).

Publications

Revue internationale avec comité de lecture

- Charles Rocabert, Carole Knibbe, Jessika Consuegra, Dominique Schneider & Guillaume Beslon (2017). « Beware batch culture: Seasonality and niche construction predicted to favor bacterial adaptive diversification ». *PLoS Computational Biology*, vol. 13, n°3, e1005459 (32 pages. doi : 10.1371/journal.pcbi.1005459.s006. HAL : hal-01508751.

Conférences internationales avec comité de lecture

- Charles Rocabert, Carole Knibbe, Jessika Consuegra, Dominique Schneider & Guillaume Beslon (2017). « Environmental seasonality drives digital populations towards stable cross-feeding ». *European Conference on Artificial Life (ECAL)*, 8 septembre 2017, Villeurbanne (France). HAL : hal-01569093.
- Charles Rocabert, Carole Knibbe, Jessika Consuegra, Dominique Schneider & Guillaume Beslon (2016). « In Silico Experimental Evolution Highlights the Influence of Environmental Seasonality on Bacterial Diversification ». *2nd EvoEvo Workshop, satellite workshop of CCS2016*, 20 septembre 2016, Amsterdam (Pays-Bas). HAL : hal-01375677.
- Jérôme Gippet, Serge Fenet, Adeline Dumet, Bernard Kaufmann & Charles Rocabert (2016). « MoRIS: Model of Routes of Invasive Spread. Human-mediated dispersal, road network and invasion parameters ». *5th International Conference on Ecology and Transportation: Integrating Transport Infrastructures with Living Landscapes*, 2 juin 2016, Lyon (France). HAL : hal-01412280.
- Charles Rocabert, Carole Knibbe & Guillaume Beslon (2015). « Towards a Integrated Evolutionary Model to Study Evolution of Evolution ». *EvoEvo Workshop (Satellite workshop of ECAL 2015)*, 24 juillet 2015, York (Royaume-Uni). HAL : hal-01252796.
- Jérôme Gippet, Charles Rocabert, Serge Fenet, Adeline Dumet & Bernard Kaufmann (2015). « Modeling and evaluating human-mediated dispersal mechanisms at landscape scale: a study of road network and invasion parameters for *Lasius neglectus* ants invasive species ». *World Conference on Natural Resource Modeling*, 29 juin 2015, Bordeaux (France). HAL : hal-01242828.

Adaptation contextuelle multi-préoccupations orientée sémantique dans le web des objets

Jury

Chantal TACONET, MC, Télécom SudParis	Rapporteuse
Pascal MOLLI, PR, Université de Nancy	Rapporteur
Frédérique LAFOREST, PR, Télécom Saint-Etienne	Examinatrice
Amélie GYRARD, Chercheur, École des Mines de Saint-Étienne	Examinatrice
Thierry MONTEIL, PR, INSA Toulouse	Examineur
Pierre-Antoine CHAMPIN, MC, Université Lyon 1	Examineur
Lionel MÉDINI, MC, Université Lyon 1	Co-encadrant

Résumé

Le Web des Objets s'inscrit dans divers domaines d'application, tels que la domotique, les entreprises, l'industrie, la médecine, la ville, et l'agriculture. Il se présente comme une couche uniforme placée au-dessus de l'Internet des Objets, afin de surmonter l'hétérogénéité des protocoles présents dans ces réseaux.

Une valeur ajoutée des applications Web des Objets est de pouvoir combiner l'accès à divers objets connectés et sources de données externes avec des techniques standards de raisonnement sémantique (RDF-S, OWL). Cela leur permet alors d'interpréter et de manipuler de ces données en tant qu'informations contextuelles. Ces informations contextuelles peuvent être exploitées par ces applications afin d'adapter leurs composants en fonction des changements dans leur environnement.

L'adaptation contextuelle est un défi majeur pour le Web des Objets. En effet, les solutions d'adaptation existantes sont soit fortement couplées avec leur domaine d'application (étant donné qu'elles reposent sur des modèles de contexte spécifiques au domaine), soit proposées comme composant logiciels autonomes, difficiles à intégrer dans des architectures Web et orientées sémantique. Cela mène alors à des problèmes d'intégration, de performance et de maintenance.

Dans cette thèse, nous proposons une solution d'adaptation contextuelle multi-préoccupations pour les applications Web des Objets, répondant à des besoins d'utilisabilité, de flexibilité, de pertinence et de performance. Notre travail se base sur un scénario pour l'agriculture numérique et se place dans le cadre de la plateforme orientée-avatar ASAWoO. Premièrement, nous proposons un méta-modèle générique permettant de concevoir des modèles contextuels standards, interopérables et réutilisables. Deuxièmement, nous présentons un cycle de vie du contexte et un workflow d'adaptation contextuelle, permettant la sémantisation de données brutes, ainsi que la contextualisation en parallèle durant l'exécution de l'application. Ce workflow combine des données issues de sources hétérogènes, telles que l'expertise du domaine, les documentations techniques des objets, les données de capteurs et de services Web, etc.

Troisièmement, nous présentons une méthode de génération de règles d'adaptations basées sur des situations contextuelles, permettant de limiter l'effort des experts et concepteurs lors de l'élaboration d'applications adaptatives. Quatrièmement, nous proposons deux optimisations pour le raisonnement contextuel : la première adapte la localisation des tâches de raisonnement en fonction du contexte, la seconde améliore le processus de maintenance incrémentale d'informations contextuelles.

Mehdi Terdjimi

Thèse

Soutenue le 18/12/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Lionel.Medini@liris.cnrs.fr

Abstract

The Web of Things (WoT) takes place in a variety of application domains (e.g. homes, enterprises, industry, healthcare, city, agriculture...). It builds a Web-based uniform layer on top of the Internet of Things (IoT) to overcome the heterogeneity of protocols present in the IoT networks.

WoT applications provide added value by combining access to connected objects and external data sources, as well as standard-based reasoning (RDF-S, OWL 2) to allow for interpretation and manipulation of gathered data as contextual information. Contextual information is then exploited to allow these applications to adapt their components to changes in their environment.

Yet, contextual adaptation is a major challenge for the WoT. Existing adaptation solutions are either tightly coupled with their application domains (as they rely on domain-specific context models) or offered as standalone software components that hardly fit in Web-based and semantic architectures. This leads to integration, performance and maintainability problems.

In this thesis, we propose a multi-purpose contextual adaptation solution for WoT applications that addresses usability, flexibility, relevance, and performance issues in such applications. Our work is based on a smart agriculture scenario running inside the avatar-based platform ASAWoO. First, we provide a generic context meta-model to build standard, interoperable et reusable context models. Second, we present a context lifecycle and a contextual adaptation workflow that provide parallel raw data semantization and contextualization at runtime, using heterogeneous sources (expert knowledge, device documentation, sensors, Web services, etc.). Third, we present a situation-driven adaptation rule design and generation at design time that eases experts and WoT application designers' work. Fourth, we provide two optimizations of contextual reasoning for the Web: the first adapts the location of reasoning tasks depending on the context, and the second improves incremental maintenance of contextual information.

Publications

Conférences internationales avec comité de lecture

- Mehdi Terdjimi, Lionel Médini, Michael Mrissa & Maria Maleshkova (2017). « Multi-purpose Adaptation in the Web of Things ». Tenth International and Interdisciplinary Conference on Modeling and Using Context, 23 juin 2017, Paris (France). HAL : hal-01462866.
- Mehdi Terdjimi, Lionel Médini, Michael Mrissa & Nicolas Le Sommer (2016). « An Avatar-Based Adaptation Workflow for the Web of Things ». 2016 IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 15 juin 2016, Paris (France). doi : 10.1109/WETICE.2016.22. HAL : hal-01308040.
- Mehdi Terdjimi, Lionel Médini & Michael Mrissa (2016). « Towards a Meta-model for Context in the Web of Things ». Karlsruhe Service Summit Workshop, 26 février 2016, Karlsruhe (Allemagne). HAL : hal-01255479.
- Mehdi Terdjimi, Lionel Médini & Michael Mrissa (2016). « HyLAR+: Improving Hybrid Location-Agnostic Reasoning with Incremental Rule-based Update ». WWW '16: 25th International World Wide Web Conference Companion, 15 avril 2016, Montréal (Canada). HAL : hal-01276558.

- Mehdi Terdjimi (2015). « Multi-level context adaptation in the Web of Things ». Doctoral Consortium at ISWC2015, 15 octobre 2015, Bethlehem (États-Unis). HAL : hal-01222490.
- Mehdi Terdjimi, Lionel Médini & Michael Mrissa (2015). « HyLAR: Hybrid Location-Agnostic Reasoning ». ESWC Developers Workshop 2015, 31 mai 2015, Portoroz (Slovénie), p. 1. HAL : hal-01154549.

Autres

- Mehdi Terdjimi & Lionel Médini (2016). « Easy client-side reasoning ». Document en anglais, présentations et journée en français. Présentation au colloque SemWeb.Pro 2016. HAL : hal-01400803.

chapitres d'ouvrages

- Lionel Médini, Michael Mrissa, Mehdi Terdjimi, El-Mehdi Khalfi, Nicolas Le Sommer, Philippe Capdepuy, Jean-Paul Jamont, Michel Ocelllo & Lionel Touseau (2017). « Building a Web of Things with Avatars ». Managing the Web of Things: Linking the Real World to the Web, Michael Sheng, Yongrui Qin, Lina Yao, and Boualem Benatallah, Morgan Kaufmann, Elsevier. doi : 10.1016/B978-0-12-809764-9.00007-X. HAL : hal-01373631.



Modèles et outils pour favoriser l'articulation entre la généricité d'un assistant intelligent et les spécificités de son usage en EIAH

Jury

Pascal LEROUX, PR, ENSIM du Mans
Sébastien IKSAL, MC, IUT de Laval
Hamid CHAACHOUA, PR, ESPE Grenoble
Viviane GUERAUD, MC, Université Grenoble Alpes
Stéphanie JEAN-DAUBIAS, PR, Université Lyon 1
Marie LEFEVRE, MC, Université Lyon 1

Rapporteur
Rapporteur
Examineur
Examinatrice
Directrice
Co-directrice

Résumé

Cette thèse en informatique se situe dans le domaine des Environnements Informatiques pour l'Apprentissage Humain (EIAH), et plus particulièrement au sein du projet AGATE (an Approach for Genericity in Assistance To complex tasks) qui vise à proposer des modèles génériques et des outils unifiés pour permettre la mise en place de systèmes d'assistance dans des applications existantes. Dans ce projet, l'éditeur d'assistance SEPIA permet aux concepteurs d'assistance de définir un système d'assistance épiphyte grâce à des règles respectant le langage aLDEAS. Ce système d'assistance peut ensuite être exécuté par le moteur d'assistance de SEPIA pour fournir de l'assistance aux utilisateurs finaux sur les applications-cibles diverses.

Dans le contexte éducatif, des enseignants souhaitent mettre en place des systèmes d'assistance pour compléter les logiciels pédagogiques ou non-pédagogiques utilisés par les apprenants. Les enseignants ont donc le rôle des concepteurs d'assistance alors que les apprenants sont des utilisateurs finaux des applications assistées. Pour permettre la mise en place d'assistance dans le contexte éducatif, nous avons identifié deux questions de recherche principales : « Quels sont les besoins en assistance informatisée dans le contexte éducatif ? » et « Comment faciliter la définition d'un système d'assistance par des enseignants non-informaticiens ? ».

Afin de répondre à la première question de recherche, nous avons étudié les applications utilisées par les enseignants au sein de leurs cours ainsi que les travaux existants qui proposent des systèmes d'assistance. Nous avons ainsi identifié des besoins d'assistance technique (utilisation de l'application, prise en main) et des besoins d'assistance pédagogique (feedbacks pédagogique, guidage pédagogique) ainsi que différents modes de déroulement d'une assistance dans le contexte éducatif.

Afin de répondre à la seconde question de recherche, nous avons confronté les modèles et outils proposés précédemment dans le projet AGATE aux besoins d'assistance identifiés dans le contexte éducatif. Les limites des modèles précédents nous ont amené à proposer deux contributions au langage aLDEAS et au système SEPIA pour les adapter au contexte éducatif.

La première limite concernait la complexité de définition de systèmes d'assistance variés en termes de déroulement. Que ce soit dans un contexte éducatif ou non, il est important de pouvoir définir facilement et de manière explicite plusieurs modes d'articulation entre les différents éléments d'un système d'assistance. Nous avons donc proposé un modèle d'articulation entre les règles aLDEAS explicitant le déroulement d'une assistance et permettant de définir des systèmes d'assistance comprenant des éléments qui se déroule de manière successive, interactive, simultanée, progressive, indépendante. Nous avons associé à ce modèle un processus d'application automatique des contraintes des modes d'articulation aux règles qui permet

Le Vinh Thai

Thèse

Soutenue le 23/06/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Stephanie.Jean-Daubias@liris.cnrs.fr

de générer et/ou modifier automatiquement les règles aLDEAS en faisant en sorte qu'elles soient toujours compatibles avec le moteur d'exécution d'assistance de SEPIA. Ce modèle et ce processus ont été implémentés dans SEPIA.

La seconde limite est propre au domaine éducatif. Elle concernait la complexité à définir un guidage pédagogique proposant un parcours entre différentes activités au sein d'une application existante. Nous avons tout d'abord proposé un modèle d'activité permettant d'identifier les activités au sein des applications. Ensuite, un modèle de guidage pédagogique permet de définir différents types de guidage pédagogique (libre, séquentiel, contextuel, temporel, personnalisé). Un patron de guidage pédagogique a été proposé pour permettre de définir un guidage pédagogique complet qui peut combiner plusieurs types de guidage pédagogique. Enfin, un processus de transformation d'un guidage pédagogique en règles aLDEAS permet de générer et modifier automatiquement les règles aLDEAS. Ce modèle d'activité, ce patron de guidage pédagogique et ce processus ont été implémentés dans SEPIA.

Publications

Conférences internationales avec comité de lecture

- Le Vinh Thai, Stéphanie Jean-Daubias, Marie Lefevre & Blandine Ginon (2016). « Model of Articulation between aLDEAS Assistance Rules ». DC CSEDU 2016, 23 avril 2016, Rome (Italie). HAL : hal-01309161.
- Le Vinh Thai, Blandine Ginon, Stéphanie Jean-Daubias & Marie Lefevre (2014). « Adapting the SEPIA system to the educational context ». EKAW, 24 novembre 2014, Linköping (Suède), pp. 1-4. HAL : hal-01301104.
- Blandine Ginon, Le Vinh Thai, Stéphanie Jean-Daubias, Marie Lefevre & Pierre-Antoine Champin (2014). « Adding epiphytic assistance systems in learning applications using the SEPIA system ». EC-TEL, 16 septembre 2014, Graz (Autriche), pp. 138-151. HAL : hal-01301072.

Conférences nationales avec comité de lecture

- Le Vinh Thai (2016). « Concevoir un guidage pédagogique à travers un système d'assistance épiphyte ». RJC EIAH, 17 juin 2016, Montpellier (France), pp. 79-80. HAL : hal-01503403.
- Le Vinh Thai, Blandine Ginon, Stéphanie Jean-Daubias, Marie Lefevre & Pierre-Antoine Champin (2015). « Modèle d'articulation entre les règles définissant un système d'assistance aLDEAS ». RJCIA, 30 juin 2015, Rennes (France). HAL : hal-01267129.
- Le-Vinh Thai & Blandine Ginon (2014). « Exploitation d'assistances épiphytes en contexte éducatif ». RJC EIAH, 16 juin 2014, La Rochelle (France), pp. 95-102. HAL : hal-01271925.

Conception et analyse d'activités collaboratives pour environnement multi-surfaces

Jury

Nadine COUTURE, PR, ESTIA
Stéphanie BUISINE, PR, EICESI Paris Nanterre
Andres LUCERO, Associate Professor, Aalto University
Olivier CHAPUIS, CR, Université Paris-Sud
Audrey SERNA, MC, INSA-Lyon
Aurélien TABARD, MC, Université Lyon 1
Sébastien GEORGE, PR, Université du Maine
Youssef AMGHAR, PR, INSA-Lyon

Rapporteur
Rapporteur
Examineur
Examineur
Co-directeur
Co-directeur
Co-directeur
Directeur de thèse

Lili Tong

Résumé

Les environnements multi-surfaces (MSE) combinent de manière flexible plusieurs surfaces interactives pour former un espace d'information continu. Les grandes surfaces, telles que les tables tactiles et les écrans muraux, sont souvent utilisées comme espace partagé pour coordonner les efforts; les surfaces mobiles, tels que tablettes et smartphones, sont souvent utilisées comme espaces personnels prenant en charge les tâches individuelles. Les environnements multi-surfaces ont prouvé leur utilité pour les activités collaboratives co-localisées, notamment pour des activités collaboratives complexes de résolution de problèmes et de prise de décision. Cependant, de nombreux défis restent ouverts, notamment liés à la diversité des environnements multi-surfaces en terme de configurations possibles, et de variabilité des dispositifs, et des activités associées. Dans ce contexte, le développement d'activités collaboratives pour environnements multi-surfaces reste complexe, et peu d'outils, à la fois théoriques et techniques en facilite le développement.

Cette dissertation étudie comment les environnements multi-surfaces peuvent soutenir la collaboration et l'apprentissage collaboratif. Elle analyse la façon dont la configuration et la forme des dispositifs façonnent les comportements collaboratifs des utilisateurs et propose des recommandations pour la conception d'activités collaboratives dans les environnements multi-surfaces. Elle offre également un outil de prototypage rapide, qui peut être facilement utilisé par des non-experts, tels que des enseignants, afin de créer des activités de prise de décision collaborative.

Abstract

Multi-surface environments (MSEs) combine several of surfaces in a variety of physical arrangements to form a seamless information space. The large surfaces, such as tabletops and wall-displays are often used as a shared space to coordinate efforts, and handheld devices, such as tablets and smartphones are regarded as personal space supporting individual tasks. MSEs have shown benefits for supporting co-located activities, especially the ones involving rich data exploration, such as complex collaborative problem-solving and decision-making activities. However, the diversity of MSE also raises questions and challenges, as different configuration and devices factors of MSE can be suited for different kinds of activities, and developing collaborative activities in MSE remains complex.

This dissertation studies how MSE can support users' collaboration in general, and collaborative learning specifically. It provides insights on how the configuration and form factors of devices in MSE shape users' collaborative behaviors, and offers implications on designing collaborative activities in

Thèse

Soutenue le 05/05/2017

Établissement :
INSA Lyon

Contact :
Youssef.Amghar@liris.cnrs.fr

MSE. It also contributes with a rapid prototyping tool, which can be easily used by non-experts, such as teachers, to create collaborative decision-making activities in MSE.

Publications

Conférences internationales avec comité de lecture

- Lili Tong, Aurélien Tabard, Sébastien George & Audrey Serna (2017). « Horizontal vs. Vertical: How the Orientation of a Large Interactive Surface Impacts Collaboration in Multi-Surface Environments ». 16th IFIP Conference on Human-Computer Interaction (INTERACT), 29 septembre 2017, Mumbai (Inde), pp. 202-222. doi : 10.1007/978-3-319-67687-6_14. HAL : hal-01597913.
- Lili Tong, Audrey Serna, Sébastien George & Aurélien Tabard (2017). « Supporting Decision-making Activities in Multi-Surface Learning Environments ». Proceedings of the 9th International Conference on Computer Supported Education (CSEDU 2017), 21 avril 2017, Porto (Portugal), pp. 70-81. HAL : hal-01493815.
- Audrey Serna, Simon Pageaud, Lili Tong, Sébastien George & Aurélien Tabard (2016). « F-formations and Collaboration Dynamics Study for Designing Mobile Collocation ». 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI 2016), 9 septembre 2016, Florence (Italie), pp. 1138-1141. doi : 10.1145/2957265.2962656. HAL : hal-01387991.
- Lili Tong, Audrey Serna, Simon Pageaud, Sébastien George & Aurélien Tabard (2016). « It's Not How You Stand, It's How You Move: F-formations and Collaboration Dynamics in a Mobile Learning Game ». In Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16). ACM, New York, NY, USA, 6 septembre 2016, Florence (Italie), pp. 318-329. doi : 10.1145/2935334.2935343. HAL : hal-01363632.
- Lili Tong, Audrey Serna, Sébastien George, Aurélien Tabard & Gilles Brochet (2014). « Interactive Surface Composition Based on Arduino in Multi-Display Environments ». the Ninth ACM International Conference on Interactive Tabletops and Surfaces, 15 novembre 2014, Dresden (Allemagne), pp. 369-374. doi : 10.1145/2669485.2669548. HAL : hal-01313181.

Conférences nationales avec comité de lecture

- Lili Tong (2015). « Supporting Design of Interaction in Multi-surface Environments for Collaborative Learning ». 27ème conférence francophone sur l'Interaction Homme-Machine, Rencontres doctorales, 27 octobre 2015, Toulouse (France), 4 pages. HAL : hal-01500439.

Lili Tong (2014). « Toward rapid prototyping tools for designing collaborative learning activities on tabletops ». RJC EIAH, 16 juin 2014, La Rochelle (France), pp. 75-80. HAL : hal-01301059.

Capitalisation pérenne de connaissances industrielles : vers des méthodes de conception incrémentales et itératives centrées sur l'utilisateur

Carine Touré

Jury

Sébastien GEORGE, PR, Université du Maine-IUT Laval
Marie-Hélène ABEL, PR, UTC, Compiègne
Serges GARLATTI, PR, IMT Atlantique
Jean-Yves PRAX, Polia Consulting
Jean-Luc DELTOUR, Société du Canal de Provence
Christine MICHEL, MC, INSA Lyon
Jean-Charles MARTY, MC, Université de Savoie

Rapporteur
Rapporteuse
Examineur
Examineur
Invité
Directrice de thèse
Co-directeur

Résumé

Dans ce travail de recherche, nous nous intéressons à la question de la pérennité de l'usage des systèmes de gestion des connaissances (SGC) dans les entreprises. Les SGC sont ces environnements informatiques qui sont mis en place dans les entreprises pour mutualiser et construire l'expertise commune grâce aux collaborateurs. Le constat montre que, malgré la rigueur employée par les entreprises pour la mise en œuvre de ces SGC, le risque d'échec des initiatives de gestion des connaissances, notamment lié à l'acceptation de ces environnements par les utilisateurs professionnels ainsi qu'à leur usage continu et durable, reste d'actualité. La persistance et l'ampleur de ce constat dans les entreprises a motivé notre intérêt d'apporter une contribution à cette question générale de recherche. Comme propositions de réponse à cette problématique, nous avons donc 1) dégagé à partir de l'état de l'art, quatre facettes qui sont requises pour favoriser l'usage pérenne d'une plateforme gérant la connaissance; 2) proposé un modèle théorique de régulation mixte qui unifie des outils de stimulation pour l'autorégulation et des outils soutenant l'accompagnement au changement et qui permet la mise en œuvre continue des différents facteurs stimulants l'usage pérenne des SGC; 3) proposé une méthodologie de conception, adaptée à ce modèle et basée sur les concepts Agile, qui intègre une méthode d'évaluation mixte de la satisfaction et de l'usage effectif ainsi que des outils d'IHM pour l'exécution des différentes itérations de notre méthodologie; 4) implémenté la méthodologie en contexte réel, à la Société du Canal de Provence, ce qui nous a permis de tester sa faisabilité et de proposer des ajustements/recommandations génériques aux concepteurs pour son application en contexte. L'outil résultant de notre implémentation a reçu un accueil positif par les utilisateurs en termes de satisfaction et d'usages.

Abstract

In this research, we are interested in the question of sustainability of the use of knowledge management systems (KMS) in companies. KMS are those IT environments that are set up in companies to share and build common expertise through collaborators. Findings show that, despite the rigor employed by companies in the implementation of these KMS, the risk of knowledge management initiatives being unsuccessful, particularly related to the acceptance and continuous use of these environments by users remains prevalent. The persistence of this fact in companies has motivated our interest to contribute to this general research question. As contributions to this problem, we have 1) identified from the state of the art, four facets that are required to promote the perennial use of a platform managing knowledge; 2) proposed a theoretical model of mixed regulation that unifies tools for self-regulation and tools to support change, and allows the continuous implementation of the various factors that stimulate the sustainable use of CMS; 3) proposed a design methodology, adapted to this model and based

Thèse

Soutenue le 19/10/2017

Établissement :
INSA Lyon

Contact :
Christine.Michel@iris.cnrs.fr

on the Agile concepts, which incorporates a mixed evaluation methodology of satisfaction and effective use as well as CHI tools for the completion of different iterations of our methodology; 4) implemented the methodology in real context at the Société du Canal de Provence, which allowed us to test its feasibility and propose generic adjustments / recommendations to designers for its application in context. The tool resulting from our implementation was positively received by the users in terms of satisfaction and usage.

Publications

Revue internationale avec comité de lecture

- Carine Touré, Christine Michel & Jean-Charles Marty (2017). « Towards extending traditional informal learning tools in the workplace with social functionalities ». *International Journal of Learning Technology*, vol. 12, n°2, pp. 119-150. HAL : hal-01512364.

Conférences internationales avec comité de lecture

- Carine Touré, Christine Michel & Jean-Charles Marty (2017). « How to promote informal learning in the workplace? The need for incremental design methods ». 9th International Conference on Computer Supported Education (CSEDU 2017), 23 avril 2017, Porto (Portugal), pp. 220-229. doi : 10.5220/0006355502200229. ArXiv : 1709.09945. HAL : hal-01478273.
- Carine Touré, Christine Michel & Jean-Charles Marty (2015). « Refinement of Knowledge Sharing Platforms to Promote Effective Use: A Use Case ». 8th IEEE International Workshop on Knowledge Acquisition, Reuse and Evaluation (KARE 2015) in conjunction with 11th International Conference on Signal-Image Technology and Internet- Based Systems (SITIS 2015), 27 novembre 2015, Bangkok (Thaïlande), pp. 680-686. doi : 10.1109/SITIS.2015.127. HAL : hal-01263423.
- Carine Touré, Christine Michel & Jean-Charles Marty (2014). « Re-designing knowledge management systems : Towards user-centred design methods integrating information architecture ». *Knowledge Management and Information Sharing*, 21 octobre 2014, Rome (Italie), pp. 298-305. ArXiv : 1601.08032. HAL : hal-01263410.
- Carine Touré, Christine Michel & Jean-Charles Marty (2014). « What if we considered awareness for sustainable Knowledge Management ? Towards a model for self regulated knowledge management systems based on acceptance models of technologies and awareness ». *International Conference KMIS (International conference on Knowledge Management and Information Sharing)*, 21 octobre 2014, Rome (Italie), pp. 413-418. ArXiv : 1601.08030. HAL : hal-01263399.

Conférences nationales avec comité de lecture

- Carine Touré, Christine Michel & Jean-Charles Marty (2017). « Adaptation des réseaux sociaux d'entreprise pour favoriser l'apprentissage informel sur le lieu de travail ». 8ème Conférence sur les Environnements Informatiques pour l'Apprentissage Humain (EIAH 2017), 9 juin 2017, Strasbourg (France), pp. 89-100. ArXiv : 1709.09413. HAL : hal-01512375.
- Carine Touré, Christine Michel & Jean-Charles Marty (2017). « Apprentissage informel sur le lieu de travail. Comment exploiter les principes des réseaux sociaux ». ORPHEE-RDV 2017 Atelier : L'Apprenant acteur principal de son parcours tout au long de sa vie, 1 février 2017, Font-Romeu (France). HAL : hal-01562345.

Correction de biais de sélection par pondération d'importance en apprentissage supervisé

Van-Tinh Tran

Jury

Alexandre AUSSEM, PR, Université Lyon 1
Marianne CLAUSEL, MC, Université Grenoble Alpes
Elisa FROMONT, MC, Université de Saint-Etienne
Christophe GONZALES, PR, Université Paris 6
Marc SEBBAN, PR, Université de Saint-Etienne

Directeur de thèse
Examinatrice
Examinatrice
Rapporteur
Rapporteur

Résumé

Dans la théorie de l'apprentissage supervisé, l'hypothèse selon laquelle l'échantillon de d'apprentissage et de test proviennent de la même distribution de probabilité, joue un rôle crucial. Malheureusement, cette hypothèse essentielle est souvent violée en présence d'un biais de sélection. Dans ce contexte, les algorithmes d'apprentissage supervisés standards peuvent souffrir d'un biais significatif. Dans cette thèse, nous abordons le problème du biais de sélection en apprentissage supervisé en utilisant la méthode de pondération de l'importance ("importance weighting" en anglais).

Dans un premier temps, nous présentons le cadre formel de l'apprentissage supervisé et discutons des effets potentiellement néfastes du biais sur les performances prédictives. Nous étudions ensuite en détail comment les techniques de pondération de l'importance permettent, sous certaines hypothèses, de corriger le biais de sélection durant l'apprentissage de modèles génératifs et discriminants. Nous étudions enfin le potentiel des réseaux bayésiens comme outils de représentation graphique des relations d'indépendances conditionnelles entre les variables du problème et celles liées au mécanisme de sélection lui-même. Nous illustrons sur des exemples simples comment la graphe, construit avec de la connaissance experte, permet d'identifier a posteriori un sous-ensemble restreint de variables sur lesquelles « agir » pour réduire le biais.

Dans un second temps, nous accordons une attention particulière au « covariate shift », i.e. un cas particulier de biais de sélection où la distribution conditionnelle $P(y|x)$ est invariante entre l'échantillon d'apprentissage et de test. Nous proposons deux méthodes pour améliorer la pondération de l'importance en présence de covariate shift. Nous montrons d'abord que le modèle non pondéré est localement moins biaisé que le modèle pondéré sur les échantillons faiblement pondérés, puis nous proposons une première méthode combinant les modèles pondérés et non pondérés afin d'améliorer les performances prédictives dans le domaine cible. Enfin, nous étudions la relation entre le covariate shift et le problème des données manquantes dans les jeux de données de petite taille et proposons une seconde méthode qui utilise des techniques d'imputation de données manquantes pour corriger le covariate shift dans des scénarios simples mais réalistes. Ces méthodes sont validées expérimentalement sur de nombreux jeux de données.

Abstract

In the theory of supervised learning, the identical assumption, i.e. the training and the test samples are drawn from the same probability distribution, plays a crucial role. Unfortunately, this essential assumption is often violated in the presence of selection bias. Under such condition, the standard supervised learning frameworks may suffer a significant bias. In this thesis, we use the importance weighting method to address the selection bias problem in supervised learning.

Thèse

Soutenue le 11/07/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Alexandre.Aussem@iris.cnrs.fr

We first introduce the supervised learning frameworks and discuss the importance of the identical assumption. We then study the importance weighting framework for the generative and the discriminative learning under a general selection scheme and investigate the potential of Bayesian Network to encode a priori assumptions about the relationships between the variables in study, including the selection variable, and to infer the independence and the conditional independence relationships that allow the selection bias to be corrected.

We pay special attention to covariate shift, i.e. a special class of selection bias where the conditional distribution, $P(y|x)$, of the training and of the test data are the same. We propose two methods to improve the importance weighting for covariate shift. We first show that the unweighted model is locally less biased than the weighted one on the low importance instances, and then propose a method that combines them in order to improve the predictive performance in the target domain. Finally, we investigate the relationship between the covariate shift and the missing data problem for data sets with small sample sizes and study a method that uses missing data imputation techniques to correct the covariate shift in some simple but realistic scenarios.

Publications

Conférences internationales avec comité de lecture

- Van-Tinh Tran & Alex Aussem (2015). « Correcting a Class of Complete Selection Bias with External Data Based on Importance Weight Estimation ». 22nd International Conference, ICONIP 2015, 12 novembre 2015, Istanbul (Turquie). doi : 10.1007/978-3-319-26555-1_13. HAL : hal-01247394.
- Van-Tinh Tran & Alex Aussem (2015). « A Practical Approach to Reduce the Learning Bias Under Covariate Shift ». ECML PKDD 2015 - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, 11 septembre 2015, Porto (Portugal), pp 71-86. doi : 10.1007/978-3-319-23525-7_5. HAL : hal-01213965.

Catégorisation des comportements de conduite en termes de consommation en carburant : une méthode de découverte de connaissances contextuelles à partir des traces d'interactions

Jury

Patrick BRÉZILLON, PR, LIP6 CNRS, Paris 6	Rapporteur
Christophe KOLSKI, PR, LAMIH, Université de Valenciennes	Rapporteur
Sylvie DESPRES, PR, Université 13, LIMICS-INSERM	Examinatrice
Sophie MIDENET, Chercheur, GRETTIA/ Ifsttar	Examinatrice
Alain MILLE, PR, Université Lyon1, LIRIS UMR 5205 CNRS	Directeur de thèse
Hélène TATTEGRAIN, LESCOT/Ifsttar	Co-directrice de thèse

Résumé

Cette thèse propose une méthode d'ingénierie des connaissances contextuelles qui permet la modélisation et l'identification du contexte explicatif d'un critère observé. Le contexte est constitué de connaissances explicatives situées permettant une représentation élicitée valide d'un objet dans la situation visée. Ces connaissances sont découvertes généralement lors de l'observation de la réalisation de l'activité dans laquelle cet objet est impliqué. Elles sont donc difficiles à décrire en début d'analyse d'une activité. Toutefois, elles restent nécessaires pour la définition, l'explication et la compréhension efficace d'une activité selon un critère observé caractérisant cette dernière. Cette thèse propose la définition progressive du contexte satisfaisant pour expliquer un critère observé lors de l'observation d'une activité. Cette recherche mobilise les traces d'interaction de l'activité analysée, précise la notion de contexte et exploite les méthodes de fouille de données pour réaliser la catégorisation et la classification d'un critère observé en distinguant les paramètres contextuels et non contextuels. L'environnement développé sur les principes des traces d'interaction, permet d'assister la découverte du contexte explicatif par une approche interactive, à l'aide des connaissances de l'analyste, de distinguer ce qui est contexte de ce qui ne l'est pas. Nous montrons qu'il est possible de construire un contexte valide, en le « découvrant » et en le formulant sous une forme générique, telle que proposée dans la littérature. Une application de la méthode a été effectuée en situation de conduite automobile pour modéliser et identifier le contexte explicatif de la consommation en carburant. En s'appuyant sur les connaissances existantes du domaine, la validation de la méthode est effectuée en étudiant qualitativement les connaissances produites sur la consommation réelle en carburant. La méthode est validée quantitativement en appliquant les règles de classifications établies sur des données collectées de l'activité de conduite. Cette illustration de l'analyse de l'activité de conduite automobile avec la méthode de découverte de connaissances contextuelles, pour déterminer le contexte explicatif de la consommation en carburant, a été effectuée à l'Ifsttar sur des données réelles collectées lors de l'activité de conduite en situation naturelle. Les expérimentations menées montrent des résultats encourageants et permettent d'envisager l'intégration de la méthode de découverte de connaissances contextuelles dans les pratiques des analystes de l'Ifsttar.

Abstract

This thesis proposes an engineering method of contextual knowledge that allows modelling and identification of explanatory context of observed criteria. The context consists of located explanatory knowledge allowing valid representation of an object in the covered situation. This knowledge is generally elicited when observing the activity performance in which the object is involved. They are therefore difficult to describe in the beginning of activity analysis but are necessary for the definition, explanation and effective

Assitan Traore

Thèse

Soutenu le 19/01/2017

Établissement :

Université Claude Bernard Lyon1

Contact :

Alain.Mille@liris.cnrs.fr

understanding of an activity according to an observed criterion characterizing this activity. This thesis proposes a progressive definition of adequate context to explain an observed criterion during activity observation. The research mobilizes interaction traces of the analysed activity, clarify context notion and uses data mining methods for classification or categorization of an observed criterion by distinguishing contextual parameters and no contextual parameters. The developed environment, based on interaction traces principles, allows to assist explanatory context discovery by interactive approach, using context analyst knowledge. We demonstrate that it's possible to build a valid context, by discovering it and by formulating it in a generic form as proposed in literature. An application of the method was performed in driving situation to identify and model the explanatory context of fuel consumption. The method validation is performed by studying produced knowledge on fuel consumption, qualitatively by relying on existing domain knowledge and quantitatively by applying classification rules established trough data collected from driving activity. This illustration of driving activity analysis with the contextual knowledge discovery method to determine explanatory context of fuel consumption was conducted at Ifsttar on real data, collected during driving activity in natural driving situation. The led experiments show encouraging results and allows considering the integration of contextual knowledge discovery method in Ifsttar analyst practices.

Publications

Conférences invité

- A Trace Analysis Based Approach for Modeling Context Components. A. Traore, H. Tattegrain, A. Mille. CONTEXT'13, Lecture Notes in Computer Science ed. Annecy. pp. 371-380. Lecture Notes in Artificial Intelligence 8175.

Copier-Coller 3D : paramétrisation cohérente de maillages triangulaires

Jury

Florent DUPONT, PR, Université Lyon 1	Directeur de thèse
Julien TIERNY, CR CNRS, Sorbonne Universités UPMC	Directeur de thèse
Guillaume LAVOUÉ, MC, INSA-Lyon	Co-encadrant
Pooran MEMARI, CR CNRS, École Polytechnique, Paris Saclay	Co-encadrant
Dominique BECHMANN, PR, Université de Strasbourg	Rapporteur
Pierre ALLIEZ, DR, INRIA Sophia-Antipolis	Rapporteur
Nicolas RAY, CR, INRIA Nancy Grand Est	Examineur

Résumé

La paramétrisation de surface est un outil fondamental pour de nombreuses applications de géométrie numérique, car elle permet le dépliage de surfaces triangulées 3D vers des domaines géométriques plus simples, tels que le plan 2D, où les traitements géométriques peuvent être réalisés de manière plus efficace ou plus simple. Les applications usuelles de la paramétrisation de surface incluent la cartographie, le placage de texture, le morphing, le transfert de détail et le re-maillage. Pour se déplier vers le plan 2D, une surface doit avoir la topologie d'un disque, ou doit être artificiellement décomposée en un atlas de cartes ayant la topologie d'un disque. Dans le premier cas, le dépliage sera habituellement accompagné d'une importante distorsion d'aire, dans le second cas, le dépliage fera émerger des artefacts visuels en raison des discontinuités de paramétrisation sur le bord des cartes. Pour surmonter ces deux difficultés, nous proposons un algorithme efficace pour la paramétrisation globale de surfaces triangulées.

Tout d'abord, les singularités coniques sont automatiquement détectées dans des endroits visuellement significatifs ; ce processus est efficace sur le plan calculatoire et vise à détecter de tels cônes aux sommets du maillage, où des valeurs élevées de distorsion de surface peuvent être prédites avant la paramétrisation réelle.

Afin d'assurer la continuité au travers des coupes coniques résultant de la découpe du maillage, des fonctions de transition affines sont utilisées ; celles-ci sont intégrées dans un système linéaire qui vise à minimiser la distorsion angulaire.

Dans cette thèse, nous présentons également un nouvel algorithme de paramétrisation croisée qui, étant donné deux maillages triangulaires d'entrée et des points de correspondance fournis par l'utilisateur, effectue des paramétrisations topologiquement et géométriquement cohérentes. La paramétrisation cohérente simultanée des maillages est réalisée en seulement quelques secondes, en résolvant au plus quatre systèmes linéaires au sens des moindres carrés.

Nous validons les résultats des algorithmes proposés en fournissant des résultats expérimentaux étendus, en démontrant l'efficacité en temps, ainsi que la qualité, illustrée en examinant les mesures de distorsion acceptées.

L'efficacité de calcul des algorithmes présentés permet leur utilisation dans des applications interactives, où l'utilisateur peut modifier ou ajouter des singularités coniques (ou des correspondances de référence pour le pipeline de paramétrisation croisée) tout en obtenant des résultats dans des temps de fonctionnement pratiques.

Ana-Maria Vintescu

Thèse

Soutenue le 28/06/2017

Établissement :

Télécom ParisTech

Contact :

Florent.Dupont@liris.cnrs.fr



Apprentissage et lissage dans les modèles de Markov à sauts et à copules

Fei Zheng

Jury

Jean-Yves TOURNERET, ENSEEIHT Toulouse
Séverine DUBUISSON, UPMC Sorbonnes Universités
François ROUEFF, Telecom ParisTech
Stéphane DERRODE, École Centrale de Lyon
Wojciech PIECZYNSKI, Telecom SudParis

Examineur
Rapporteur
Rapporteur
Directeur de thèse
Co-directeur de thèse

Résumé

Les modèles de Markov à sauts (appelés JMS pour Jump Markov System) sont utilisés dans de nombreux domaines tels que la poursuite de cibles, le traitement des signaux sismiques et la finance, étant donné leur bonne capacité à modéliser des systèmes non-linéaires et non-gaussiens. De nombreux travaux ont étudié les modèles de Markov linéaires pour lesquels bien souvent la restauration de données est réalisée grâce à des méthodes d'échantillonnage statistique de type Markov Chain Monte-Carlo (MCMC). Dans cette thèse, nous avons cherché des solutions alternatives aux méthodes MCMC et proposons deux originalités principales. La première a consisté à proposer un algorithme de restauration non supervisée d'un JMS particulier appelé « modèle de Markov couple à sauts conditionnellement gaussiens » (noté CGPMSM). Cet algorithme combine une méthode d'estimation des paramètres basée sur le principe Espérance-Maximisation (EM) et une méthode efficace pour lisser les données à partir des paramètres estimés. La deuxième originalité a consisté à étendre un CGPMSM spécifique appelé CGOMSM par l'introduction des copules. Ce modèle, appelé GCOMSM, permet de considérer des distributions plus générales que les distributions gaussiennes tout en conservant des méthodes de restauration optimales et rapides. Nous avons équipé ce modèle d'une méthode d'estimation des paramètres appelée GICE-LS, combinant le principe de la méthode d'estimation conditionnelle itérative généralisée et le principe des moindres carrés linéaires. Toutes les méthodes sont évaluées sur des données simulées. En particulier, les performances de GCOMSM sont discutées au regard de modèles de Markov non-linéaires et non-gaussiens tels que la volatilité stochastique, très utilisée dans le domaine de la finance.

Abstract

Switching Markov Models, also called Jump Markov Systems (JMS), are widely used in many fields such as target tracking, seismic signal processing and finance, since they can approach non-Gaussian non-linear systems. A considerable amount of related work studies linear JMS in which data restoration is achieved by Markov Chain Monte-Carlo (MCMC) methods. In this dissertation, we try to find alternative restoration solution for JMS to MCMC methods. The main contribution of our work includes two parts. Firstly, an algorithm of unsupervised restoration for a recent linear JMS known as Conditionally Gaussian Pairwise Markov Switching Model (CGPMSM) is proposed. This algorithm combines a parameter estimation method named Double EM, which is based on the Expectation-Maximization (EM) principle applied twice sequentially, and an efficient approach for smoothing with estimated parameters. Secondly, we extend a specific sub-model of CGPMSM known as Conditionally Gaussian Observed Markov Switching Model (CGOMSM) to a more general one, named Generalized Conditionally Observed Markov Switching Model (GCOMSM) by introducing Copulas. Comparing to CGOMSM, the proposed GCOMSM adopts inherently more flexible distributions and non-linear structures, while optimal restoration is

Thèse

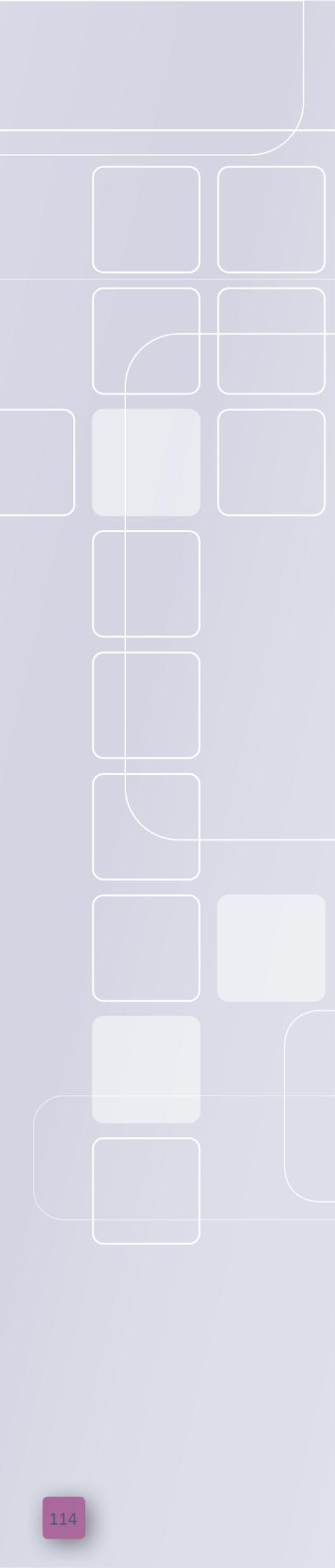
Soutenu le 18/12/2017

Établissement :

Ecole Centrale de Lyon

Contact :

Stephane.Derrode@liris.cnrs.fr



feasible. In addition, an identification method called GICE-LS based on the Generalized Iterative Conditional Estimation (GICE) and the Least-Square (LS) principles is proposed for GCOMSM to approximate any non-Gaussian non-linear systems from their sample data set. All proposed methods are tested by simulation. Moreover, the performance of GCOMSM is discussed by application on other generable non-Gaussian non-linear Markov models, for example, on stochastic volatility models which are of great importance in finance.

Laboratoire d'Informatique en Image et Systèmes d'information

Contact

Université Claude Bernard Lyon 1

Bâtiment Nautibus

25 avenue Pierre de Coubertin

F-69622 Villeurbanne Cedex

☎ +33 472 43 36 10

✉ secretariat@liris.cnrs.fr

<http://liris.cnrs.fr/>

