

Sujet : Apprentissage multi-label partiel de l'expérience patient à partir de texte

Contexte et objectifs

Nous constatons aujourd'hui une généralisation de la collecte de commentaires textuels non structurés provenant des patients des prestataires de soins de santé. Cependant, les techniques d'analyse permettant d'examiner de manière approfondie ces retours demeurent peu appliquées dans ce domaine. Afin de combler cette lacune, nous proposons un stage visant à mettre en pratique des approches novatrices de vectorisations de données textuelles et d'annotation multi-labels corrective à partir d'un vaste ensemble de commentaires émis par les patients à propos de leurs établissements de soins. L'objectif est de nourrir et enrichir les démarches d'amélioration de la qualité en se basant sur les ressentis et l'expérience des patients, une priorité pour les établissements de santé de haut niveau.

L'objectif de ce stage est de mettre en place un framework novateur, incluant (1) l'extraction des informations significatives à partir de commentaires textuels libres, (2) l'annotation correctives multi-labels et (2) la détermination de facteurs pertinents expliquant ainsi cette annotation. Les perceptions de la qualité des services de santé étant traditionnellement exprimées sous forme d'évaluations numériques, cette initiative permettra d'apporter une compréhension approfondie des aspects qualitatifs qui ne sont pas capturés par les chiffres.

Les principales tâches du stage seront les suivantes :

- État de l'art sur l'analyse de l'expérience du patient et l'utilisation des approches d'intelligence artificielle dans ce domaine novateur.
- Etat de l'art sur les techniques de traitement automatique de langage et l'apprentissage multi-labels partiel en vue de leur utilisation sur les données collectées et la correction de annotations induites.
- Investigation dans les modèles d'apprentissage automatiques neuronaux, voire profonds, et étude de leur comportement sur les données collectées.
- Utilisation de techniques d'analyse statistique pour évaluer la qualité des modèles développés et pour identifier les tendances dans les données collectées.

Références

- KHANBHAI, Mustafa, ANYADI, Patrick, SYMONS, Joshua, *et al.* Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. *BMJ Health & Care Informatics*, 2021, vol. 28, no 1.
- MINAEE, Shervin, KALCHBRENNER, Nal, CAMBRIA, Erik, *et al.* Deep learning--based text classification: a comprehensive review. *ACM computing surveys (CSUR)*, 2021, vol. 54, no 3, p. 1-40.
- Ming-Kun Xie, Sheng-Jun Huang. **Partial Multi-Label Learning With Noisy Label Identification**. *IEEE Trans. Pattern Anal. Mach. Intell.* 44(7): 3676-3687 (2022)

Encadrants

- Khalid Benabdeslem (Univ-Lyon1, LIRIS) : khalid.benabdeslem@univ-lyon1.fr
- Abdel Malek Habi (HCL) : ext-abdelmalek.habi@chu-lyon.fr