

Détection des humeurs dépressives sur les réseaux sociaux chinois à partir d'une combinaison de plongements lexicaux et de méthodes textométriques

Jinyuan Xu¹, Pierre Magistry², Mathieu Valette³

¹Ertim, Inalco – jinyuan.xu@inalco.fr

²Ertim, Inalco – pierre.magistry@inalco.fr

³Ertim, Inalco – mathieu.valette@inalco.fr

Abstract

Studies aimed at detecting users at risk of depression on Chinese social media generally rely on metadata or limited textual characteristics of the users, as well as large language models, foregoing any semantic analysis and ignoring the linguistic expression of emotions. We propose an architecture for extracting and analyzing semantic features based on textual statistics to address this issue. Our corpus is derived from the Sina Weibo Depression Dataset. We will use textual statistics methods to analyze the corpus with the goal of constructing an interpretable semantic model, to which we will then combine machine learning and deep learning methods. The results will be analyzed to determine the contribution of the semantic model to detection and its interpretability.

Keywords : detection, classification, specificity calculation, repeated segments, semantic model, machine learning, deep learning, depression, Weibo, Chinese

Résumé

Les études pour détecter les utilisateurs à risque de dépression sur les réseaux sociaux chinois se basent généralement sur les métadonnées ou les caractéristiques textuelles limitées des utilisateurs et de grands modèles de langues (LLMs). Elles s'affranchissent de toute analyse sémantique et négligent l'expression linguistique des émotions. Nous proposons une architecture d'extraction et d'analyse des caractéristiques sémantiques basée sur les statistiques textuelles pour remédier à ce problème. Notre corpus est extrait du jeu de données *Sina Weibo Depression Dataset*. Nous utiliserons les méthodes de statistiques textuelles pour l'analyse du corpus avec l'objectif d'en construire un modèle sémantique interprétable à partir duquel nous combinons ensuite des méthodes d'apprentissage automatique et d'apprentissage profond. Les résultats seront analysés pour déterminer l'apport du modèle sémantique à la détection et son interprétabilité.

Mots clés : détection, classification, calcul de spécificités, segments répétés, modèle sémantique, apprentissage automatique, apprentissage profond, dépression, Weibo, chinois