Institut national
des langues
et civilisations orientales

Yizhou XU

Détection d'anomalies dans les textes pour la veille

Directeur de thèse :
Frédérique SEGOND

Date de soutenance :
22 janvier 2025

Résumé

Le travail présenté dans cette thèse, mené au sein de Chapsvision dans le cadre d'une convention CIFRE, a pour objectif de développer et affiner des méthodes pour identifier des anomalies dans les textes, avec une attention particulière portée aux scénarios de veille stratégique. La détection d'anomalies, essentielle pour identifier erreurs, changements critiques ou activités suspectes, a été peu explorée dans le contexte des données textuelles.

Pour combler cette lacune, notre approche repose sur deux axes. D'une part, des algorithmes d'apprentissage automatique conçus pour d'autres types de données (séries temporelles, données tabulaires, images) ont été adaptés pour traiter les spécificités des textes via des méthodes semi-supervisées, non supervisées et faiblement supervisées. D'autre part, nous avons intégré les modèles les grands modèles de langue (LLMs) dans notre solutionpour la détection d'anomalies, en expérimentant avec diverses stratégies de conception de prompts.

Les contributions de cette recherche permettent d'avancer l'état de l'art dans le domaine de la détection d'anomalies dans les données textuelles, offrant des perspectives prometteuses pour des applications en veille stratégique et au-delà.

Signé électroniquement par
Frédérique SEGOND
Le 22/11/2024 à 12:36