

Résumé :

Cette thèse s'inscrit dans le cadre d'apprentissage artificiel pour les applications du Texte Mining. Nous avons proposé, en général, des nouvelles approches relatives au progrès des applications d'analyse des sentiments comme les sentiments commerciaux des clients d'amazone, de la classification automatique et de la catégorisation multiple des textes, en effet les approches que nous avons proposées se focalisent sur l'amélioration des processus d'extraction des caractéristiques, y compris la vectorisation des textes et la gestion de la grande dimensionnalité des descripteurs. Notre première contribution encadre les mécanismes nécessaires pour réussir la classification automatique, où nous avons comparé entre différents systèmes de classification, ainsi l'étude comparative que nous avons réalisée, a prouvé qu'un meilleur choix des méthodes de racinisation, de vectorisation, et l'hybridation des algorithmes d'apprentissage, influencent fortement la performance et la fiabilité de ces systèmes. La deuxième contribution porte sur la proposition d'une nouvelle approche de vectorisation probabiliste pour rendre la catégorisation neuronale des textes faisable et performante. Aussi dans notre troisième intervention, nous proposons une amélioration notable sur les descripteurs flous, où nous suggérons l'automatisation des systèmes d'inférences flous améliorant la sélection des règles d'inférence. Cette proposition renforce la vectorisation floue des textes et progresse le rendement des classifieurs automatiques. Notre dernière contribution, présenté dans cette manuscrite, exploite l'amélioration de la logique floue par la logique neutrosophique afin de présenter un nouveau modèle neutrosophique avancé pour la rénovation de la vectorisation floue des données textuelles. Ainsi, la mise en œuvre d'un nouveau descripteur vectoriel qui se base sur la philosophie neutrosophique pour représenter les termes pertinents des bases des données textuelles de références, permet d'avoir des systèmes de catégorisation des textes robustes avec des performances intéressantes. Certes, l'ensemble des outils, les comparaisons et les résultats des nouvelles contributions seront affichés et discutés au sein de cette mémoire.

Mots clés :

Text Mining, Apprentissage Artificiel, Extraction des caractéristiques, Vectorisation des Textes, Logique Floue, Logique Neutrosophique, Descripteurs Neutrosophique, Classification multiple, Analyse des sentiments, Hybridation des classifieurs.

STUDY OF UNSTRUCTURED DATA ANALYSIS AND CATEGORIZATION SYSTEMS BASED ON MACHINE LEARNING: TEXT MINING APPLICATIONS.

Abstract:

This thesis is part of the artificial learning framework for Text Mining applications. In general, we have proposed new approaches that enhance the performance of sentiment analysis applications such as amazon customers' commercial sentiments, automatic classification, and multiple categorizations of texts. Indeed, the suggested approaches focus on improving feature extraction processes, including text vectorization, and handling high dimensionality of descriptors. Thus, our first contribution outlines the mechanisms required for a successful text organization, where we compared different text classification systems. Hence, the comparative study showed that a good choice of the stemming methods, vectorization, and the learning algorithms combination strongly influence the performance and the reliability of these systems. The second contribution introduces a new text vectorization approach based on probabilistic reasoning. The proposal allows making an efficient neural texts categorization system. In our third intervention, we propose a significant improvement on fuzzy descriptors, where we suggest the automation of fuzzy inference systems improving the selection of inference rules. This proposal boosts the fuzzy texts vectorization and the performance of automatic classifiers. Ensuing, we use the neutrosophic logic to present a new advanced neutrosophic model. The proposed approach replaces the fuzzy term representation and practices the neutrosophic inference system to calculate relevant term weight. Therefore, using the new neutrosophic descriptor for text classification applications shows satisfying results. In this thesis, the tools and the results of the discussed contributions will be displayed and discussed.

Key Words:

Text Mining, Artificial Learning, Features Extraction, Text Vectorization, Fuzzy Logic, Neutrosophic Logic, Neutrosophic Descriptor, Multiple Classification, Sentiment Analysis, Hybrid Classifiers.