

Stage M2 Recherche en Informatique :
**Ajout de connaissances linguistiques pour des
tâches de classification fondées sur les motifs
séquentiels**

Responsables : Mathieu Roche*, Pascal Poncelet**, Maguelonne Teisseire***

*Équipe TAL, LIRMM,

mroche@lirmm.fr

**LGI2P / LIRMM

Pascal.Poncelet@ema.fr

***Équipe Tatoo / LIRMM

teisseire@lirmm.fr

Résumé. Proposition et évaluation de méthodes d'ajout de connaissances linguistiques aux tâches de classification fondées sur les motifs séquentiels.

1 Description

De nombreuses tâches de classification s'appuient sur des méthodes statistiques particulièrement bien adaptées pour le traitement de gros volumes de données textuelles. Les approches fondées sur les motifs séquentiels se sont révélées pertinentes pour la classification de textes (Jaillet et al. (2006)). Les travaux menés jusqu'à maintenant n'intègrent pas de connaissances linguistiques pour améliorer les méthodes mises an œuvre. Les travaux de Moreau et al. (2007) ont par exemple montré que trois types de connaissances peuvent être ajoutées et combinées : les informations lexicales, syntaxiques et sémantiques.

- Au cours du stage, un état de l'art complet sur les types de connaissances linguistiques et les différentes approches pour leur prise en compte dans les méthodes de classification sera étudié.

Stage M2 Recherche en Informatique

- Une approche fondée sur l’ajout de connaissances lexicales et syntaxiques aux méthodes de classification fondées sur les motifs séquentiels sera proposée.
- La méthode proposée sera évaluée sur différents jeux de données de domaines et de langues différents.

Références

- Jaillet, S., A. Laurent, et M. Teisseire (2006). Sequential Patterns for Text Categorization. *International Journal of Intelligent Data Analysis (IDA)* 10(3).
- Moreau, F., V. Claveau, et P. Sébillot (2007). Intégrer plus de connaissances linguistiques en recherche d’information peut-il augmenter les performances des systèmes ? In *Actes de la 4ème Conférence en Recherche d’Informations et Applications, CORIA 07, Saint-Etienne, France*, pp. 223–238.

Summary

Adding linguistic knowledge to the classification tasks based on Sequential Patterns.