

Un système de vote pour la classification de textes d'opinion

Michel Plantié*, Mathieu Roche**, Gérard Dray*

* LGI2P, Ecole des Mines d'Alès, Site EERIE
(michel.plantie, gerard.dray)@ema.fr

** LIRMM, UMR 5506, Univ. Montpellier 2, CNRS,
mathieu.roche@lirmm.fr

Résumé : Les tâches de classification textuelle ont souvent pour objectif de regrouper thématiquement différents textes. Dans cet article, nous nous sommes intéressés à la classification de documents en fonction des opinions et jugements de valeurs qu'ils contiennent. L'approche proposée est fondée sur un système de vote utilisant plusieurs méthodes de classification.

1 Introduction

La classification de textes a pour objectif le regroupement de documents selon différents critères. Dans les travaux présentés dans cet article, nous nous intéressons à la classification de textes d'opinion qui consiste à classer les textes selon un jugement tel que l'aspect positif ou négatif d'une critique, l'aspect favorable ou défavorable donné par un expert, etc. Nous proposons dans cet article une approche fondée sur plusieurs classifieurs combinés à un système de vote. Dans un premier temps, nous présentons les corpus du défi DEFT'07 (Grouin *et al.*, 2007) sur lesquels nous avons mené nos expérimentations ainsi que les représentations des textes utilisées. La section 3 décrit les classifieurs et les systèmes de vote proposés. Enfin, la partie 4 présente les résultats obtenus.

2 Représentation des données textuelles

La troisième édition du défi francophone DEFT'07 consistait à déterminer des catégories de jugements à partir de quatre corpus français très différents en terme de thème, taille, tournures de phrases, richesse du vocabulaire, représentation des catégories de jugement :

- ✓ **Corpus 1** : Critiques de films, livres, spectacles et bandes dessinées. Trois catégories : bon, moyen, mauvais.
- ✓ **Corpus 2** : Critiques de jeux vidéo. Trois catégories : bon, moyen, mauvais.
- ✓ **Corpus 3** : Commentaires de révision d'articles de conférences scientifiques. Trois catégories : acceptation, acceptation sous conditions, rejet.