

# Combining Document-Level Topic Dependent and Topic Independent Evidences for Opinion Retrieval

Malik Muhammad Saad Missen<sup>1</sup>, Faiza Belbachir<sup>2</sup> and Guillaume Cabanac<sup>2</sup>

<sup>1</sup> Dept. of Computer Science and IT,  
The Islamia University of Bahawalpur, Pakistan  
[Saad.Malik@iub.edu.pk](mailto:Saad.Malik@iub.edu.pk)

<sup>2</sup> Université de Toulouse  
IRIT UMR 5505 CNRS  
Toulouse, France  
[{Belbachir, Guillaume.Cabanac}@irit.fr](mailto:{Belbachir, Guillaume.Cabanac}@irit.fr)

## **Résumé**

Dans cet article, nous proposons une approche pour la détection d'opinions qui combine la dépendance du sujet par rapport au document et l'indépendance du sujet. L'apport de notre approche réside dans la simplicité et l'efficacité des caractéristiques (features) utilisées. Les expérimentations sur la collection de données TREC Blog06 montrent que notre approche donne une amélioration significative par rapport à ce que TREC fournit comme meilleurs résultats du «baseline» pour 50 sujets de l'année 2007. Pour confirmer l'efficacité de notre approche, une autre expérimentation est faite sur les sujets de l'année 2006 et 2008. Nous avons également comparé et analysé les résultats de notre approche en utilisant différentes formulations des caractéristiques proposées. La comparaison des résultats avec les approches précédentes (la dépendance ou l'indépendance du sujet) révèle que notre approche donne les meilleurs résultats enregistrés (à notre connaissance) pour les sujets de l'année 2007 en utilisant deux de ces cinq «baseline». Une analyse des résultats révèle que les sujets les plus populaires sont susceptibles de contenir plus d'opinions que les moins populaires. En expérimentant avec différents étiquetages grammaticaux (c.-à-d. verbes, adjectifs et adverbes), nous constatons que la combinaison des adjectifs et des adverbes pourrait être un meilleur indicateur de la subjectivité d'un document que d'autres combinaisons d'étiquetages grammaticaux.

**Mots-clés** : détection des opinions, Blogs, Sujets, Subjectivité, Sentiment

## **Abstract**

In this paper, we propose an opinion-finding approach that combines document-level topic dependent and topic independent evidences. The uniqueness of our approach lies in simplicity and effectiveness of the

evidences being used. The experimentations with TREC Blog06 data collection show that our approach gives a significant improvement over TREC provided strong baselines for 50 topics of year 2007. Further experimentation using the TREC strongest baseline is performed to verify effectiveness of our approach for topics of years 2006 and 2008. In addition to this, we also compare and analyze the results of our approach using different formulations of proposed features. A comparison of results with previous approaches (whether topic-dependent or topic-independent) reveals that our approach produces the best results ever reported (to the best of our knowledge) for topics of year 2007 using two of these five baselines. An analysis of the results reveals that more popular topics are likely to be more opinionated than the less popular ones. By experimenting with different parts-of-speeches (i.e., verbs, adjectives and adverbs), we find out that combining adjectives and adverbs could be a better indicator of subjectivity of a document than other combinations of parts-of-speeches.

**Keywords:** Opinion Detection, Blogs, Topic, Subjectivity, Sentiment, TREC.