

# ARC '16

مؤتمر مؤسسة قطر  
السنوي للبحوث  
QATAR FOUNDATION  
ANNUAL RESEARCH  
CONFERENCE



Towards World-class  
Research and Innovation

## Information Communications Technology Pillar

<http://dx.doi.org/10.5339/qfarc.2016.ICTPP3059>

### **Sentiment Analysis in Comments Associated to News Articles: Application to Al Jazeera Comments**

**Khalid Al-Kubaisi, Abdelaali Hassaine, Ali Jaoua**

Qatar University, QA

Email: 198700982@qu.edu.qa

Sentiment analysis is a very important research task that aims at understanding the general sentiment of a specific community or group of people. Sentiment analysis of Arabic content is still in its early development stages. In the scope of Islamic content mining, sentiment analysis helps understanding what topics Muslims around the world are discussing, which topics are trending and also which topics will be trending in the future.

This study has been conducted on a dataset of 5000 comments on news articles collected from Al Jazeera Arabic website. All articles were about the recent war against the Islamic State. The database has been annotated using Crowdfunder which is website for crowdsourcing annotations of datasets. Users manually selected whether the sentiment associated with the comment was positive or negative or neutral. Each comment has been annotated by four different users and each annotation is associated with a confidence level between 0 and 1. The confidence level corresponds to whether the users who annotated the same comment agreed or not (1 corresponds to full agreement between the four annotators and 0 to full disagreement).

Our method represents the corpus by a binary relation between the set of comments ( $x$ ) and the set of words ( $y$ ). A relation exists between the comment ( $x$ ) and the word ( $y$ ) if, and only if, ( $x$ ) contains ( $y$ ). Three binary relations are created for comments associated with positive, negative and neutral sentiments. Our method then extracts keywords from the obtained binary relations using the hyper concept method [1]. This method decomposes the original relation into non-overlapping rectangles and highlights for each rectangle the most representative keyword. The output is a list of keywords sorted in a hierarchical ordering of importance. The obtained keyword list associated with positive, negative and neutral comments are fed into a random forest classifier of 1000 random trees in order to predict the sentiment associated with each comment of the test set.

**Cite this article as:** Al-Kubaisi K, Hassaine A, Jaoua A. (2016). Sentiment Analysis in Comments Associated to News Articles: Application to Al Jazeera Comments. Qatar Foundation Annual Research Conference Proceedings 2016: ICTPP3059 <http://dx.doi.org/10.5339/qfarc.2016.ICTPP3059>.

Experiments have been conducted after splitting the database into 70% training and 30% testing subsets. Our method achieves a correct classification rate of 71% when considering annotations with all values of confidence and even 89% when only considering the annotation with a confidence value equal to 1. These results are very promising and testify of the relevance of the extracted keywords.

In conclusion, the hyper concept method extracts discriminative keywords which are used in order to successfully distinguish between comments containing positive, negative and neutral sentiments. Future work includes performing further experiments by using a varying threshold level for the confidence value. Moreover, by applying a part of speech tagger, it is planned to perform keyword extraction on words corresponding to specific grammatical roles (adjectives, verbs, nouns... etc.). Finally, it is also planned to test this method on publicly available datasets such as the Rotten Tomatoes Movie Reviews dataset [2].

### Acknowledgment

This contribution was made possible by NPRP grant #06-1220-1-233 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.

### References

- [1] A. Hassaine, S. Mecheter, and A. Jaoua. "Text Categorization Using Hyper Rectangular Keyword Extraction: Application to News Articles Classification." *Relational and Algebraic Methods in Computer Science*. Springer International Publishing, 2015. 312–325.
- [2] B. Pang and L. Lee. 2005. "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". In *ACL*, pages 115–124.