Sentiment Analysis in a Cross-Media Analysis Framework

Yonas Woldemariam
Department of Computing Science
Umeå University
Umeå, Sweden
Email: yonasd@cs.umu.se

***Abstract*—This paper describes the implementation and integration of a sentiment analysis pipeline into the ongoing open source cross-media analysis framework. As a part of the integration, we evaluate and compare two broad categories of sentiment analysis methods, namely lexicon-based and machine learning (ML). We explore which method is appropriate to detect sentiments from forum discussion posts. Our evaluation shows that the ML model outperforms the lexicon-based by 9.88% accuracy on variable length positive, negative, and neutral comments. However, the lexicon-based shows better performance on classifying positive comments. We also obtained that the F1-score by the Lexicon-based is greater by 0.16 from the ML.**

I. INTRODUCTION

There is a massive increase of multimedia data on the Internet due to rapidly growing ubiquitous web access. However, analyzing those raw media resources to discover their hidden semantics is a challenging task. So, improving the searchability of the multimedia contents on the web is one of the most strong demands, especially for online audio/video content providers. This problem motivated the developers of the ongoing FP7 EU research project called Media in Context (MICO, http://www.mico-project.eu). MICO aims at providing a cross-media analysis framework, including orchestrated chain analysis components to extract semantics from the media resources in a cross-media context.

We are concerned with the implementation and integration of text analysis modules into the MICO platform, including sentiment analysis component. Sentiment analysis deals with the task of opinion mining from a text. In general, sentiment analysis methods are classified into lexicon-based (Musto et al, 2014) and machine learning-based (Vinodhini & Chandrasekaran, 2012;Socker et al, 2013). Machine learning methods make use of learning algorithms and classifier models trained on a known dataset. The lexicon-based approach involves calculating sentiment polarity using dictionary of words annotated with sentiment scores.

We compare these two broad categories of sentiment analysis methods regarding their prediction accuracy and study which method outperforms the other. We chose our test case to be the Zooniverse (https://www.zooniverse.org) forum discussion domain, as it is one of the show cases of MICO. Our focus is to run sentiment analysis on the texts extracted from the forum discussions to assess what the users think about the quality of the images posted on the Zooniverse site and generally about their services. Unlike the comments found in social media such as Twitter, the nature of the texts we get from the Zooniverse Snapshot Serengeti is highly characterized by the descriptions about observed images rather than explicit opinions. Thus, studying the sentiment analysis with such kind of texts creates new research challenges due to their unique features. We run the lexicon-based and Recursive Neural Tensor Network (RNTN) models (Socker et al, 2013). Our study shows that RNTN outperforms lexicon-based by 9.88% accuracy.

The paper has been published in the Proceedings of IEEE International Conference on Big Data Analysis (ICBDA), pages 1-5, Hangzhou, China

II. RELATED LITERATURE REVIEW

There are some comparative studies (Maharani, 2013;Padmaja et al, 2014) on lexicon-based versus machine learning approaches. In (Maharani, 2013), the Twitter testing dataset with a total of 1,000 tweets have been used. The result shows that machine learning methods produce better accuracy rate than lexicon-based approach. As the authors in (Maharani, 2013) stated, the significant influence from the lexical database has been set as a reference in determining positive and negative opinions. In (Padmaja et al, 2014), 1,675 sentences have been used from political news domain, among the methods the best F-measure shown by support vector machine.

III. AN OVERVIEW OF THE SENTIMENT ANALYSIS COMPONENT IN THE MICO ARCHITECTURE

The MICO framework follows a distributed service-oriented architectural approach(illustrated in Figure 1.), where analysis components run independently and share communication and persistence infrastructure (Schaffert & Fernandez, 2014) .

The core services provided by the framework include media analysis, search, and recommendation. Once the analysis components get registered with the framework and up running, the user can load a media with its context. Depending on the request made by the user, the orchestration service set up a workflow of extractors and trigger them according to the execution plan built during the request. Finally, the resulting

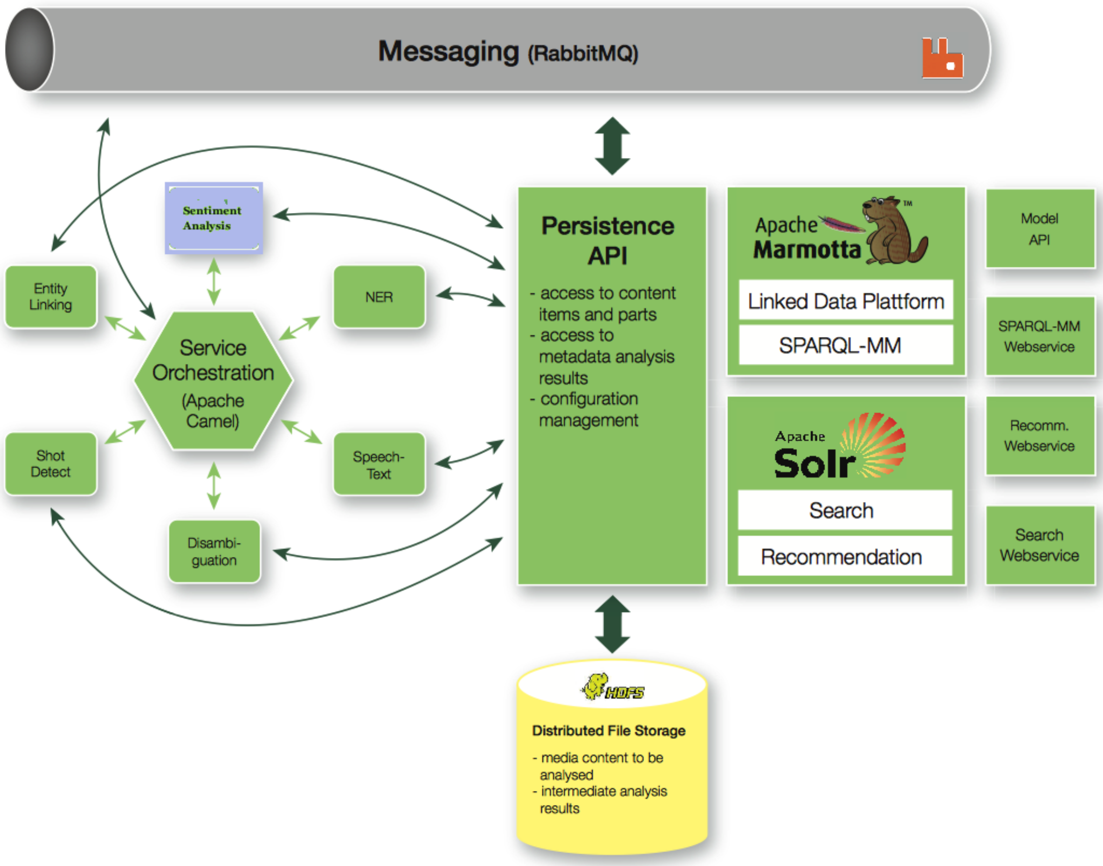

Fig. 1. MICO General Architecture, adopted from (Schaffert & Fernandez, 2014).

output from this collaborative analysis contains a bit of information added by each extractor, for instance, the language of a text.

The input for the sentiment analysis component is a set of documents, pre-processed by the chat-room cleaner module, which removes non-standard characters and repeated spaces, and produces plain text. Then the NLP sub-component does tokenization, stemming, split into sentences, and other useful text processing tasks. Then the sentiment analysis service annotates the resulting text with a sentiment label.

IV. EXPERIMENTAL EVALUATION AND DISCUSSION

We randomly chose 600 sample tweets from the Zooniverse as a test dataset. We apply commonly used performance metrics (Olson & Delen, 2008) in sentiment analysis. These are Accuracy (A), Precision (P), Recall (R) and $F1-score$. The corresponding equation for each metric is given below:

$$A = AI/T \tag{1}$$

$$P = TP/(TP + FP) \tag{2}$$

$$R = TP/(TP + FN) \tag{3}$$

$$F1 - score = 2(PR)/(P + R) \tag{4}$$

Where AI is the number of accurately predicted comments, T is the total number of comments, TP is the number of accurately predicted positive comments, FP is the number of incorrectly predicted as positive comments and FN is the number of positive comments, but incorrectly predicted as negative comments.

Metrics	Lexicon-based	RNTN
Accuracy	38.45	48.34
Precision	0.63	0.82
Recall	0.96	0.46
F-score	0.74	0.59

TABLE I
EVALUATION RESULTS OF THE LEXICON-BASED AND RNTN.

Our experimental evaluation is shown in Table I, the RNTN method outperforms the lexicon-based by 9.88%. However, the lexicon-based shows better performance on the positive comments. The lexicon-based also scores 0.96 value of R, and that means every positive instance(which does not include the reversed negative instance e.g. “not bad”) is correctly classified. Even if a wide gap has been shown by the two methods in terms of P and R, they have quite closer F1-score value. We also observed that stronger sentiment often builds up in the longer phrases and the majority of the shorter phrases are neutral, which supports the claim demonstrated in (2012;Socker et al, 2013). Some of the comments are hard to classify even by human due to their vagueness and ambiguity.

Another interesting fact is that, unlike to the lexicon-based algorithm, the RNTN seems to be powerful to capture negation and learn the sentiment of phrases following the contrastive conjunction “but”. In the case of lexicon-based, the major reason for the prediction errors is the algorithm fails to understand the context of the words including negation. In the case of RNTN, the prediction errors are caused by the mismatching of domain knowledge between training dataset and test dataset. The training dataset is collected from the movie reviews whereas the test dataset is obtained from citizen-science domain, Zooniverse. As a result, the algorithm gets challenged to recognize some unseen positive/negative phrases specific to the domain. Therefore, the straightforward approach to improve the prediction accuracy is to further train the RNTN model on Snapshot Serngeti data.

V. ACKNOWLEDGMENTS

We acknowledge the financial support from the EU FP7 MICO project. We also thank the Zooniverse people for providing the dataset used in this study. Finally, We thank Adil Blessing (student of Umeå University), for labeling the test dataset.

REFERENCES

C. Musto, G. Semeraro, and M. Polignano, 2014. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. *Proceedings of the 8th International Workshop on Information Filtering and Retrieval*, pages 59-68, Pisa, Italy.

G. Vinodhini and R. Chandrasekaran, 2012. Sentiment analysis and opinion mining: A survey. *Journal of the Advanced Research in Computer Science and Software Engineering*, 2(6):283-292. .

L. Olson, and Delen, 2008. Advanced data mining techniques. Springer.

R. Socher, A. Perelygin, J. Wu, J. Chuang, C. Manning, A. Ng, and C. Potts, 2013. Recursive deep models for semantic compositionality over a sentiment treebank. *Proceedings of Empirical Methods in Natural Language Processing*, pages 1631-1642,Seattle, Washington, USA.

S. Padmaja, S. Sameen, and S. Bandu, 2014. Evaluating sentiment analysis methods and identifying scope of negation in newspaper articles. *Journal of Advanced Research in Artificial Intelligence*, 3(11):1-6

S. Schaffert and S.Fernandez, 2014. D6.1.1 mico system architecture and development guide. MICO.

W. Maharani, 2013. Microblogging sentiment analysis with lexical based and machine learning approaches. *Proceedings in the Information and Communication Technology (ICoICT)*, pages 439-443, Bandung, Indonesia.