

Negation handling for Amharic sentiment classification

Girma Neshir¹, Andreas Rauber² and Solomon Atnafu³

¹Addis Ababa University, IT Doctoral Program, AAU, Ethiopia,

²Technical University of Vienna, Institute of Information Systems Engineering, Austria,

³Addis Ababa University, Department of Computer Science, Ethiopia

¹girma.neshir@aau.edu.et, ²rauber@ifs.tuwien.ac.at and ³solomon.atnafu@aau.edu.et

Abstract

User generated content is bringing new aspects of processing data on the web. Due to the advancement of World Wide Web technology, users are not only consumer of web contents but also they are producers of contents in the form of text, audio, video and picture. This study focuses on the analysis of textual contents with subjective information (referring to sentiment analysis). Most of conventional approaches of sentiment analysis do not effectively capture negation in languages where there are limited computational linguistic resources (e.g. Amharic). For this research, we proposed Amharic negation handling framework for Amharic sentiment classification. The proposed framework combines the lexicon based sentiment classification approach and character ngram based machine learning algorithms including LR(logistic regression) and NB(Naïve Bayesian). Finally, the performance of framework is evaluated using the annotated Amharic news comments. The system is performing the best of all models and the baselines with accuracy of 98.0. The result is compared with the baselines (without negation handling and word level ngram model).

Keywords: Negation handling algorithm, Amharic sentiment classification, Amharic sentiment lexicon, char level, word level ngram, machine learning, hybrid

Introduction

Users usually express their feelings, emotions and opinions as comments in response to the posted news, photo, audio and video. Currently, opinionated sources are increasing in languages other than English. However, Amharic sentiment analysis researches are very few as it has no sufficient linguistic resources for linguistic preprocessing and sentiment analysis. There are several challenges in lexicon based sentiment analysis. One of these is that handling negation in the text. The most common approach for negation handling is carried out relying on negation keywords. However, it is complex to identify the scope of negation

where the process of correctly identifying the part of the text affected by the presence of negation word. Negation handling in Amharic language has not been studied previously. Thus, this research develops an automatic method to handle negation and combined with char ngram features for Amharic sentiment classification relying on these linguistic features. The research questions to be addressed in this work are as follows: (a) how can we automatically detect negation words in Amharic texts? (b) how can we design a framework for handling negation in Amharic sentiment analysis? (c) how to capture char level ngram features for improving Amharic sentiment analysis in Social media(e.g. Facebook) and (d) how can we measure the performance of the proposed framework?

Related works

In this section, we briefly present the key related works. Asmi et. al. Asmi & Ishaya (2012) proposed an approach that can detect negation and considers scope of negation relying on syntactic dependency for sentiment analysis. In Amalia et al. (2018), proposed rule based negation handling and its scope based on syntactic parsing of Indonesian language for machine learning based sentiment analysis in twitter and the result of support vector machine performs well as compared with other experiments. The F-Score on two Twitter data sets is improved by 1.79% and 2.69% from the existing baseline without negation handling. In Farooq et al. (2017), develop negation scope handling strategies relying on the linguistic features of the language. The accuracy of negation scope identification is 83.3% which outperforms very well above the baseline. In Diamantini et al. (2016), develops negation scope detection relying on dependency parsing tree and semantic disambiguation technique and evaluated on integrated social networks in real time. The performance of negation handling is outperforming with accuracy of 6% more than the baseline. The work in Heerschop et al. (2011) develops nega-

comments. We evaluated the usefulness of the combination of Negation Handling (NH) and character level ngram based machine learning models for sentiment classification of Amharic facebook news comments. We call the combination (i.e. hybrid) of rule based NH and machine learning algorithms (logistic regression and Naïve Bayesian) using char ngram based tfidf features for Amharic sentiment classification. The proposed approaches are evaluated by measuring accuracy of individual and their combinations for Amharic text sentiment classification. Amharic negation scope identification and handling is recommended for further researches. We also suggest method to consider char ngram embedding features from corpus of the same domain(e.g. Facebook news comments).

References

Rizkiana Amalia, Moch Arif Bijaksana, and Dhinta Darmantoro. Negation handling in sentiment classification using rule-based adapted from indonesian language syntactic for indonesian text in twitter. In *Journal of Physics: Conference Series*, volume 971, page 012039. IOP Publishing, 2011.

Amna Asmi and Tanko Ishaya. Negation identification and calculation in sentiment analysis. In *The second international conference on advances in information mining and management*, pages 1--7, 2012.

Claudia Diamantini, Alex Mircoli, and Domenico Potena. A negation handling technique for sentiment analysis. In *2016 International Conference on Collaboration Technologies and Systems (CTS)*, pages 188--195. IEEE, 2016.

Martine Enger, Erik Velldal, and Lilja Øvrelid. An open-source tool for negation detection: a maximum-margin approach. In *Proceedings of the Workshop Computational Semantics Beyond Events and Roles*, pages 64--69, 2017.

Umar Farooq, Hasan Mansoor, Antoine Nongaillard, Yacine Ouzrout, and Muhammad Abdul Qadir. Negation handling in sentiment analysis at sentence level. *JCP*, 12(5):470--478, 2017.

Bas Heerschop, Paul van Iterson, Alexander Hogenboom, Flavius Frasinca, and Uzay Kaymak. Accounting for negation in sentiment analysis. In *11th DutchBelgian Information Retrieval Workshop (DIR 2011)*, pages 38--39, 2011.

Girma Neshir Alemneh, Andreas Rauber, and Solomon Atnafu. Dictionary Based Amharic Sentiment Lexicon Generation, pages 311--326. 08 2019.