Sentiment Analysis Model for Opinionated Awngi Text: Case of Music Reviews

Melese Mihret
University of Gondar/Gondar, Ethiopia
melesemihret42@gmail.com

Muluneh Atinaf
Addis Ababa University/Addis Ababa, Ethiopia
mulunehatinaf@yahoo.com

Abstract

The analysis of sentiments is imperative to make a decision for individuals, organizations, and governments. Due to the rapid growth of Awngi (Agew) text on the web, there is no available corpus annotated for sentiment analysis. In this paper, we present a SA model for the Awngi language spoken in Ethiopia, by using supervised machine learning approach. We developed our corpus by collecting around 1500 posts from online sources. This research is begun to build and evaluate the model for opinionated Awngi music reviews. Thus, pre-processing techniques have been employed to clean the data, to convert transliterations to the native Ethiopic script for accessibility and convenience to typing and to change the words to their base form by removing the inflectional morphemes. After pre-processing, the corpus is manually annotated by three the language professional for giving polarity, and rate, their level of confidence in their selection and sentiment intensity scale values. To improve the calculation method of feature selection and weighting and proposed a more suitable SA algorithm for feature extraction named CHI and weight calculation named TF IDF, increasing the proportion and weight of sentiment words in the feature words. We employed Support Vector Machines (SVM), Naïve Bayes (NB) and Maximum Entropy (MxEn) machine learning algorithms. Generally, the results are encouraging, despite the morphological challenge in Awngi, the data cleanness and small size of data. We are believed that the results could improve further with a larger corpus.

1 Motivation

Sentiment analysis (SA) can be applied to any human language; some approaches could be language specific. Most SA studies have been conducted in the English language and the methods cannot be directly implemented in other languages. Awngi language in the Cushitic family that is rich with morphology. The root-template morphology, which characterizes Cushitic languages results in high word productivity and necessitates NLP pre-processing techniques so as to analyze aspect-based sentiment correctly (Gebremeskel, 2010).

According to (Andrea, Ferri, & Grifoni, 2015; Chunping, Liu, Zhang, & Yang, 2016; Asghar, Khan, Ahmad, & Kundi, 2014) both DL and SL analysis is useful, but on this levels analysis do not discover detail features and couldn't address what exactly people liked or not. In this research, the study focuses on feature-level sentiment classification using a machine learning method for Awngi music reviews.

As stated in (Ge, Pengbo, & Yongquan, 2015; Sarawgi, 2017) studying opinionated text, mainly in FL, is more challenging because it involves a fine-grained sentiment analysis. Mostly, research on FL sentiment analysis often deals with data in English, while data from other languages are less explored. In Ethiopia, research on this area is still at the beginning, especially for local languages like the Awngi language. This research adds an existing task and an effort is being made to create resources and techniques to be used in this task. Generally, a study on this level used to extract detail features from opinion holders suggests what, they want to address.

2 Approaches

This research work relies on using machine learning approaches and techniques to resolve the opinionated Awngi music reviews. To come up with an effective analysis, the proposed model consists of: (pre-processing, feature extraction, determine opinion words, collecting opinions, generating a structured summary, and make a decision for features).

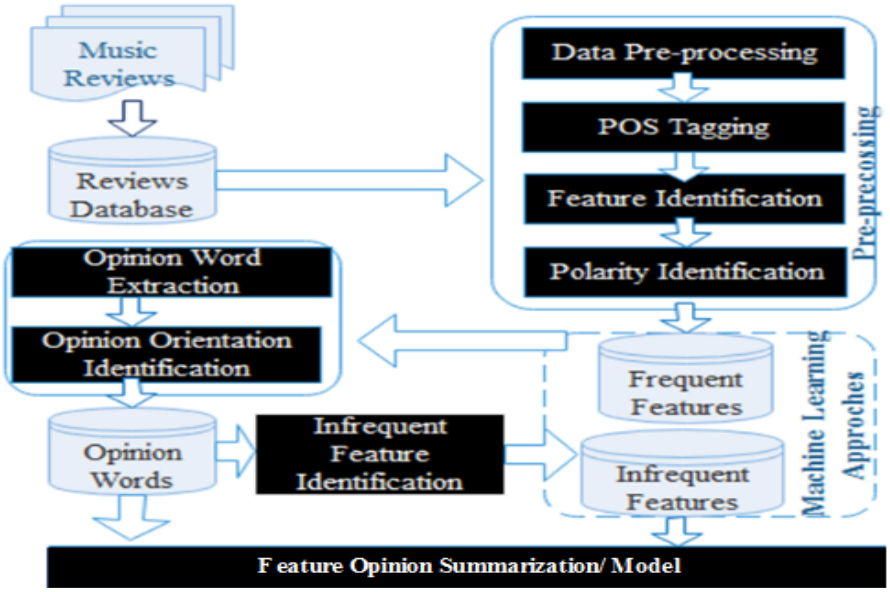

Figure 1: The Proposed Machine Learning Sentiment Analysis Model for Awngi Text.

In this study for opinion feature extraction and polarity determinations, we select positive and negative reviews from the corpus: Extracting features (Nouns/noun phrases from the reviews), Opinion words (adjectives), determine their polarities (positive and negative and neutral) and then calculate sentiment scores for each review.

3 Experiments

The experiment is done to measure the overall performance of the developed supervised machine learning sentiment analysis model. From the available corpus, the portion of the training set is 1125 posts and the remaining 375 posts are set aside as test data set. This data set contains comments that have been labeled positive, negative and neutral sentiments which contain 865 positive reviews, 393 negative reviews, and 242 neutral reviews. Each post in the test set is used as an input post one by one, and the system returns their polarity. Because of the three learning algorithms was presented with sentiment words (Unigrams and Bigrams) as a feature. We achieved an average accuracy of 75% NB, 70% MxEn and 79 % SVM performances were obtained.

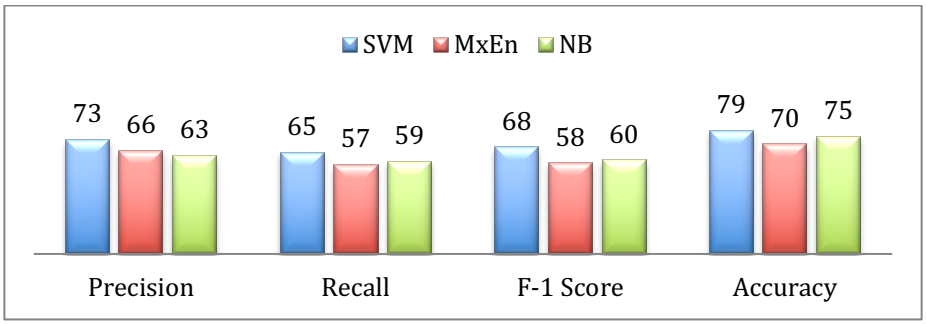

Figure 2: Describes the comparison result of NB, MxEn and SVM with precision, Recall, F-1 Score and Accuracy

To our knowledge, this work is very useful for Awngi language research and usability and on sentiment analysis in particular. We strongly believe that we have identified a research gap with this work, and developed different pre-processing steps for the task. Our contributions are twofold: we have built an annotated targeted sentiment corpus for the music reviews, and a developed a SA model for the Awngi language. It is our belief that we rigorously follow the standard machine learning and evaluation approaches to sentiment analysis. We conclude with challenges for future work so as to come up with a better solution.

Reference

Andrea, A. D., Ferri, F., & Grifoni, P. (2015). Approaches, Tools and Applications for Sentiment Analysis, Implementation, *125* (3), 26–33.

Asghar, M. Z., Khan, A., Ahmad, S., & Kundi, F. M. (2014). A Review of Feature Extraction in Sentiment Analysis. *Journal of Basic and Applied Scientific Research*, *4* (3), 181–186.

Chunping, O., Liu, Y., Zhang, S., & Yang, X. (2016). Features-level Sentiment Analysis of Movie Reviews, *81* (CST), 110–113. https://doi.org/10.14257/astl.2015.81.23

Ge, W., Pengbo, P., & Yongquan, L. (2015). Feature Extraction and Opinion Summarization in Chinese Reviews. *The Open Automation and Control Systems Journal*, *7* (1), 533–539. https://doi.org/10.2174/1874444301507010533

Gebremeskel, S. (2010). School of Graduate Studies Sentiment Mining Model for Opinionated By: Selama Gebremeskel Faculty of Computer and Mathematical Sciences Sentiment Mining Model for Opinionated Amharic Texts.

Sarawgi, K. (2017). Opinion Mining : Aspect Level Sentiment Analysis using SentiWordNet and Amazon Web Services, *158* (6), 31–36.