Multi-Lingual Sentiment Classification

Amal El Fallah Seghrouchni, Fatima zahra Salmam

Key-words: sentiment analysis, machine learning, natural language processing.

1   Research Motivation

Sentiment analysis or opinion mining is a task that aims to find out what people think about a given subject, specific entities, or online product. It can be performed on document level [1], sentence-level[2], or aspect-level [3], [4]. Multinational companies operating in several markets often use it. These companies are interested in consumers' opinions about their products and services in different countries (different languages). Therefore, multi-lingual sentiment analysis can help companies to better understand their internal and external environment and anticipate related events. They can thus improve and optimize the effectiveness of their marketing and sales actions through strategic segmentation or by implementing a personalized marketing strategy for each type of customer.

Sentiment analysis from textual sources represents a set of methods that are based on Natural Language Processing (NLP) techniques, it aims to build a classifier from text to determine whether a text is positive, negative, or neutral by extracting particular words or phrases.

There are two main categories of analysis: lexical analysis and automatic analysis. The lexical analysis approach consists in finding the sentiment via a semantic analysis of the words. It relies on classifying the sentence based on existing sentence examples for which sentiment has already been identified. In contrast, the automatic approach is based on machine learning techniques. The sentiment analysis task is usually represented as a classification problem in which a classifier is fed with text and returns the corresponding category, e.g. positive, negative or neutral.

2   Problem Statement

Sentiment analysis faces many difficulties due to the complexity of human language. Indeed, it is difficult to believe that a machine can be trained to understand irony, sarcasm, grammatical variations, slang, or social and cultural specificities that can be expressed in an online comment.

In addition, languages evolve, and vocabulary is constantly changing; all these elements must be taken into account during the analysis. For example, in the sentence "There is a bus strike today, so I will have to take a cab to go to work. Great!", the word Great can be interpreted as a positive sentiment, which is not the case.

Another difficulty with sentiment analysis is that there is no predefined structure for a comment. Comments are free-form, and are often full of many useless details that are irrelevant for the analysis, and must be identified and eliminated. Furthermore, a comment can refer to several

aspects of the same product or service and express different sentiments for each of them. For example: "the hotel was very nice but the service leaves a lot to be desired". In these complex cases, it is difficult to characterize the comment using only one value. It will be necessary to divide the comment into two or more parts, and then analyze the sentiment corresponding to each part.

On the other hand, most research [5], [6], [7] in the field of multi-Lingual sentiment classification uses machine translation from minor languages to English (and then performs sentiment analysis for English), which leads to losing the meaning of sentences.

3 Reasearch scope

The objective is to contribute to the problem of multi-lingual sentiment classification and propose a language-independent approach that could provide a good level of classification accuracy in several languages without using machine translations [8]. The approach must deal with not well-formed sentences in terms of Natural Language grammar, structure, formality, missing punctuations, and other texting disfluencies.

4 Admission Criteria

The PhD position is proposed by the International Center of Artificial Intelligence of Morocco, of the Mohammed VI Polytechnic University.

Applicants with excellent cursus must be holders of a Master's, an engineering or an equivalent recognized degree in Computer Science. In addition, they should have skills in Programming (Python and C++) and good communication skills in English. Particular attention will be given to the suitability of this research project with the applicant's background.

References

[1] Abinash Tripathy, Abhishek Anand, and Santanu Kumar Rath. "Document-level sentiment classification using hybrid machine learning approach". In: *Knowledge and Information Systems* 53.3 (2017), pp. 805–831.

[2] Bo Pang and Lillian Lee. "A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts". In: *arXiv preprint cs/0409058* (2004).

[3] Ambreen Nazir et al. "Issues and challenges of aspect-based sentiment analysis: a comprehensive survey". In: *IEEE Transactions on Affective Computing* (2020).

[4] Kalim Sattar et al. "A Multi-Layer Network for Aspect-based Cross-lingual Sentiment Classification". In: *IEEE Access* (2021).

[5] Xinjie Zhou, Xiaojun Wan, and Jianguo Xiao. "CLOpinionMiner: Opinion target extraction in a cross-language scenario". In: *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 23.4 (2015), pp. 619–630.

[6] Patrik Lambert. "Aspect-level cross-lingual sentiment classification with constrained SMT". In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 2015, pp. 781–787.

[7] Alexis Conneau et al. "Word translation without parallel data". In: *arXiv preprint arXiv:1710.04087* (2017).

[8]   Erfan Ghadery et al. "Mncn: A multilingual ngram-based convolutional network for aspect category detection in online reviews". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. 01. 2019, pp. 6441–6448.