1 Introduction

In recent years, companies are demanding finer market analysis in order to increase their revenues, improve their products or services, and carry out better marketing campaigns. Most current proposals perform polarity analysis on a message level, that is, they cannot analyse the opinion regarding the different attributes that are referenced in a message. An attribute is a feature, an aspect, or a component of the entity to which a message refers. Attribute-based polarity analysis computes the polarity of a message regarding each of the attributes that are referenced in that message.

Current approaches focuses on using ontologies to give more importance to an attribute respect to the domain [1]. Other proposals explore the neighbours of a message in order to enrich them with sentiment and semantic distance distributions using an unsupervised lexicon-based method [2].

In this paper, we introduce a novel attribute-based polarity analysis proposal that explores conditions. Our model is based on semantic skeletons that carry the information that our proposal extracts from the messages, including attributes, their values and conditions. Our hypothesis is that semantic skeletons and the opinion about them can be extracted from actual messages using syntactic patterns, which are regular expressions that use the part-of-speech tags.
2 Research Proposal

Our method focuses on extracting so-called semantic skeletons from messages that are gathered from social media. We assume that every message is related to a so-called item, which refers to a product or a service on which a user is commenting. Semantic skeletons are represented as tuples of the following form: \((d, c, a, v)\), here \(d\) is a describable, that is, an item or a component of an item, \(c\) is a condition, \(a\) is an attribute, and \(v\) is a value.

Our method consists of the following steps: preprocessing, intended to clean the messages so as to facilitate performing the following steps; extracting; summarising, in order to create cluster to facilitate the understanding of what the people are saying about the item; and analysing, to compute whether the polarity of a summarised cluster of semantic skeletons is positive, neutral or negative using a lexicon-based implementation provided by Opileak [3].

The extracting step extracts semantic skeletons from messages using a set of rules. A rule to extract semantic skeletons is represented as a tuple of the following form: \((p, t)\), where \(p\) is a standard Java regular expression to which we refer to as the syntactic pattern and \(t\) is a replacement template to which we refer to as the semantic skeleton template. In addition to attributes, we need to extract conditions and the item. We think that we might also extract the conditions using syntactic patterns, but we did not validate this hypothesis yet. We are planning on exploring Conditional Random Fields to discover the boundaries of conditional sentences [5].

We have to design an evaluation plan to the whole proposal, that is, given a set of messages that refers to an item, evaluate, using performance measures from confusion matrix, the polarity calculated over each cluster of semantic skeleton. For now, we focus on evaluation of semantic skeletons extraction, that consist on building a set of syntactic patterns using an own tool developed for this purpose [4]. We perform the dataset creation on a set of reviews (more than 450,000 messages) from different websites like Ciao, Amazon, or Dooyoo. We use a subdataset of 10,413 messages that have any matches with a set of experimental syntactic patterns and a test set of labelled messages. These patterns have a precision of 0.70.

3 Conclusions

In this paper, we have presented a new method for attribute-based polarity analysis that has the novel of extracting conditions. This method exploits the POS-tagging information in order to make easy the extraction via syntactic patterns that we use to define the templates of the extracted information. This information is summarised in order to group similar semantic skeletons and finally, we calculate the polarity with a lexicon-based algorithm. Now we are in the phase of building the set of syntactic patterns that give us the best results.