SMACk: An Argumentation Framework for Opinion Mining

Mauro Dragoni, Célia da Costa Pereira, Andrea G. B. Tettamanzi, Serena Villata

SMACk: An Argumentation Framework for Opinion Mining

Mauro Dragoni[1], **Célia da Costa Pereira**[2], **Andrea G.B. Tettamanzi**[2], **Serena Villata**[3]
[1]Fondazione Bruno Kessler, Trento, Italy
[2]Université Nice Sophia Antipolis, I3S, UMR 7271, Sophia Antipolis, France
[3]CNRS, I3S, UMR 7271, Sophia Antipolis, France
dragoni@fbk.eu celia.pereira—andrea.tettamanzi@unice.fr villata@i3s.unice.fr

Abstract

The extraction of the relevant and debated opinions from online social media and commercial websites is an emerging task in the opinion mining research field. Its growing relevance is mainly due to the impact of exploiting such techniques in different application domains from social science analysis to personal advertising. In this demo, we present our opinion summary application built on top of an argumentation framework, a standard AI framework whose value is to exchange, communicate and resolve possibly conflicting viewpoints in distributed scenarios. We show how our application is able to extract relevant and debated opinions from a set of documents containing user-generated content from online commercial websites.

1 Introduction

Argumentation theory is a reasoning model based on the construction and evaluation of information pieces called *arguments*. Arguments are supposed to support, contradict, explain statements, and they are used to support decision making. Argumentation theory involves different ways for analyzing arguments and their relationships. A famous framework is the one called *abstract argumentation* proposed by [Rahwan and Simari, 2009], which sees each argument as an abstract entity and in which arguments are related to each other by means of attack relations. What distinguishes argumentation-based discussions from other approaches is that proposals can be supported by the arguments that justify, or oppose, them. This permits greater flexibility than in other decision-making and communication schemes since, for instance, it makes it possible to persuade the other actors involved in the discussion to change their view of a claim by identifying information or knowledge that is not being considered, or by introducing a new relevant factor in the middle of a negotiation or to resolve an impasse [Rahwan and Simari, 2009].

For all these reasons, the aim of applying argumentation theory to the opinion mining field is twofold. On the one hand, we want to build an argumentation graph by extracting, from opinion content, triples structured as $\langle Aspect, Opinion, Attack\rangle$ or $\langle Aspect, Opinion, Support\rangle$, where $Aspect$ is a property of an object (for instance, the "display" of a monitor), and $Opinion$ is a value connected with such a property in the interval $[0, 1]$, where 0 represents the most negative polarity ("bad" or "very bad", depending on the granularity of the scale of judgment), 1 is the most positive polarity ("good" or "very good"), and values inbetween represent graded or mixed opinions. On the other hand, we want to exploit implicit information users provide supporting their viewpoint by adopting quick tagging facilities (for example, the "like" and "dislike" buttons of some Web interfaces) to integrate the generated argumentation framework with information that is not explicitly mentioned in the text, but that we are able to infer through the analysis of users' actions. For instance, if it is extracted from a product review the triple $\langle$*"display"*, *"good"*, $Support\rangle$, and a user "likes" such review, it means that all the nodes that will be inserted in the argumentation graph will have a further $Support$ information even if it has not been explicitly mentioned in textual form.

In this demo, we present SMACk 1.0, our argumentation-based opinion mining framework which is focused on the analysis of online user-generated content. More precisely, such a framework is based on the use of abstract bipolar argumentation theory [Rahwan and Simari, 2009] supporting the detection and extraction of relevant opinions from a set of textual documents.

The paper is organized as follows: in Section 2, we discuss why this framework is important in the opinion mining field, and in particular, we briefly discuss why the presented solution is preferable with respect to the adoption of text summarization approaches. Finally, in Section 3, we describe the content of the demonstration of SMACk.

2 The Context

In the field of opinion mining [Kumar and Ravi, 2015] different solutions have been proposed in the literature, in particular concerning aspect-based opinion mining where the aim is not to compute the polarity of an entire text, but to infer the polarity associated with single aspects mentioned in a document [Dragoni *et al.*, 2015]. However, this task is limited to the analysis of single documents. If we extend the application of opinion mining techniques to a repository of documents, open challenges are the detection of the most relevant

aspects mentioned by users, how they discuss them, and the implementation of a scalable approach for managing a huge amount of documents.

The extraction of relevant information from a text is an activity usually performed by text summarization systems [Nenkova and McKeown, 2012]. Text summarization aims to extract key information from a text, which is then presented to users as summaries or, in some cases, as lists of relevant keywords. The extraction task can be done on a single document or on a set thereof. The main advantage of applying text summarization techniques is the possibility to relieve users from reading entire documents containing irrelevant details with respect to the topic of the document and the final goal of users.

While text summarization is the most suitable solution when parts of the documents are not particularly informative from the user perspective, they are not a solution when the goal is to go beyond the extraction of interesting information from texts. In the scenario proposed in our demo, the adoption of text summarization approaches would not be the most suitable solution for the following reasons. First of all, the considered documents already contain only informative information. Indeed, generally, when a user writes a review, she does not spend a lot of time and space to contextualize the review, but she provides only the most important aspects she wants to share with the community. Therefore, it is not requested to summarize documents content, but to aggregate the different opinions they express. Second, the desired output of the analysis of user-generated content is not only the extraction of the most relevant aspects, but the detection of the most debated ones about a particular topic. The approach has to be able to extract all the aspects of the given topic and to rank them from several perspectives (user agreement, user disagreement, polarities, etc.).

Our demo presents an application which aims to go a step ahead with respect to text summarization by inferring, for a given topic, which are the most debated aspects and which are, instead, the aspects users agreed about.

3 SMACk Demonstration

The SMACk framework can be applied to several contexts with different levels of complexity. The main examples are in the social science domain, where a huge amount of text needs to be analyzed for detecting the mood of people with respect to different debated topics, or the analysis of online user-generated content about products or services. The latter is the use case we selected for demonstrating our framework. In particular, we consider a set of product reviews belonging to one of the categories used into the Amazon website (for this showcase, we will use the "Electronics" category). We analyzed a set of $50,000$ reviews extracted from the Dranziera[1] dataset.

Construction of the Argumentation Graph The first task that will be presented in the demonstration is the construction of the argumentation graph given a set of user-generated contents about the same kind of product (for instance, "laptops"). For each text, the set of aspects and the qualities associated with them are extracted and presented to the user under the form of structured information. In this way, the user is able to have a preliminary view about which are the most important aspects that are taken into account by users for judging the product she is interested in. In particular, we will show:

- a graph containing a node for each aspect that is detected in the text, and an edge connecting the different aspects only when they are mentioned in the same review. Therefore, each edge represents the correlation strength between two aspects;
- the size of each node, from a graphical point of view, is proportional to the number of mentions that the aspect represented by the node obtained in the texts;
- the color of each node is proportional to the average polarity of the aspect represented by the node that has been inferred from the text.

Computing the argument acceptability After the construction of the argumentation graph, the second task is to show how the application of argumentation theory ranking-based semantics helps in the identification of the most interesting aspects from the user perspective. The computation of the acceptability degree of the arguments is presented as an animation that will show how the graph changes at each iteration of the algorithm. At the end, the application will show the result of the analysis by presenting to the user a list of the most impactful arguments and their degree of importance.

This kind of analysis, declined in this specific use case, enables the extraction of relevant information that can be used for different purposes. For example:

- detecting which are the weaker aspects of a product in order to drive future actions on possible improvements;
- the exploitation of weak and strong aspects may inspire the development of personalized advertisement tools where information about the user viewpoints may help in making advertisement campaigns more effective;
- supporting users in making more informative decisions without the need to dig in the huge amount of (possibly technical) reviews about a product.

[1] http://goo.gl/7jK4Rp

References

[Dragoni *et al.*, 2015] Mauro Dragoni, Andrea G.B. Tettamanzi, and Célia da Costa Pereira. Propagating and aggregating fuzzy polarities for concept-level sentiment analysis. *Cognitive Computation*, 7(2):186–197, 2015.

[Kumar and Ravi, 2015] P. Ravi Kumar and Vadlamani Ravi. A survey on opinion mining and sentiment analysis: Tasks, approaches and applications. *Knowl.-Based Syst.*, 89:14–46, 2015.

[Nenkova and McKeown, 2012] Ani Nenkova and Kathleen McKeown. A survey of text summarization techniques. In *Mining Text Data*, pages 43–76. 2012.

[Rahwan and Simari, 2009] Iyad Rahwan and Guillermo R. Simari. *Argumentation in Artificial Intelligence*. Springer Publishing Company, Incorporated, 1st edition, 2009.