Pattern Recognition Letters 000 (2017) 1–9

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships

Teresa Alsinet, Josep Argelich, Ramón Béjar*, Cèsar Fernández, Carles Mateu, Jordi Planes

Artificial Intelligence Group – INSPIRES Research Centre, Universitat de Lleida, C/Jaume II 69, Lleida 25001, Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Argumentation reasoning
Probabilistic valued relationships
Support propagation
Twitter discussions

ABSTRACT

Twitter is one of the most widely used social networks when it comes to sharing and criticizing relevant news and events. In order to understand the major opinions accepted and rejected in different domains by Twitter users, in a recent work we developed an analysis system based on valued abstract argumentation to model and reason about the social acceptance of tweets, considering different information sources from the social network. Given a Twitter discussion, the system outputs the set of accepted tweets from the discussion, considering two kinds of relationship between tweets: criticism and support. In this paper, we introduce and investigate a natural extension of the system, in which relationships between tweets are associated with a probability value, indicating the uncertainty that the relationships hold. An important element in our system is the notion of an uncertainty threshold, which characterizes how much uncertainty on probability values we are willing to tolerate: given an uncertainty threshold α, we reject criticism and support relationships with probability below α. We also extend our analysis system by incorporating support propagation when computing the social relevance of tweets. To this end, we extend the abstract argumentation framework with a new valuation function that propagates the support between tweets by taking into account not only the social relevance of tweets but also the probability that the support relationship holds, provided that it is above the specified uncertainty threshold α. In order to test these new extensions, we analyze different Twitter discussions from the political domain. Our analysis shows that the social support of the accepted tweets is typically much stronger than the one for the rejected tweets. Also, the set of accepted tweets seems to be very stable with respect to changes to the social support of the tweets, and therefore even when considering support propagation we mainly observe differences in such set when using the more permissive probability thresholds.

1. Motivation and antecedents

Since its inception, in early 2006, Twitter has become one of the fastest-growing and most influential social networks. What started as a simple service to post quick and short, up to 140-character-long, status updates, has grown into one of the keystones of social debate, even being used to promote and organize action, or to empower people politically [30,39].

For instance, when it comes to politics and social issues, Twitter has either been involved in or has helped to create debate, ranging from legislation debate, as in the case of the #TTIP treaty debate, to #guncontrol debates, Wikileaks and Snowden leaks; debates on social unrest, as in #ocuppywallstreet and #spanishrevolution [22]; or even revolutions and protests, such as the Egyptian and Tunisian revolts in the Arab Spring, also called the "Twitter Revolutions" [27,41], Iranian election protests in 2009, or the Tiananmen commemoration protests in Hong Kong [31]. From all these cases, it can be seen that the usage of Twitter is not only a status publishing tool (its original intended use), but rather it also serves as an announcement and information dissemination tool, and as a forum-like discussion media, the most interesting use to our study.

In order to understand the major opinions accepted and rejected in different domains by Twitter users, in a recent work we developed a system for analysis of discussions in Twitter [2]. The system architecture has two main components: a discussion retrieval and a reasoning system. The discussion retrieval component allows us to move from a discussion in Twitter (a set of tweets) in natural language to a specialized structure modeled as a weighted graph, which is computed taking into account two semantic relationships between tweets: criticism and support, and three different attributes of a tweet: the number of followers of the author, the number of retweets and the number of favorites. The reasoning system component maps the weighted graph into a valued

* Corresponding author.
E-mail address: ramon@diei.udl.cat (R. Béjar).

http://dx.doi.org/10.1016/j.patrec.2017.07.004
0167-8655/

argumentation framework and the set of socially accepted tweets in the discussion is evaluated and computed from the weights or values assigned to the tweets in the discussion and the criticism relationships between them.

In this paper we introduce and investigate a natural extension of the system in which relationships between tweets are associated with a probability value, indicating the uncertainty that the relationships hold, and support relationships are propagated between tweets, reinforcing the set of socially accepted tweets in a discussion. In fact, when constructing relationships between tweets from informal descriptions expressed in natural language with other attributes such as emoticons, jargon, onomatopoeia and abbreviations, it is often evident that there is uncertainty about whether some of the criticism and support relationships hold. An important element of our system is the notion of an uncertainty threshold, which characterizes how much uncertainty on probability values we are prepared to tolerate: given an uncertainty threshold α, we would be prepared to disregard criticism and support relationships up to α. We therefore obtain a valued abstract argumentation framework where arguments are tweets, argument values are the weights used to model the relative social relevance of tweets from data obtained from Twitter, and attacks between arguments denote criticism relationships between tweets whose probability of fulfillment is greater than or equal to α. In order to reinforce the set of socially accepted tweets in a discussion, in this work we also propose to extend the system by propagating support relationships between tweets. To this end, we extend the valued abstract argumentation framework with a new valuation function that propagates the support between tweets by taking into account not only the weight of tweets, but also the probability that the support relationship holds, provided that it is above a specified cut-off level α.

We test our system by analyzing the effect of the uncertainty on relationships, the probability thresholds, and the support propagation on different Twitter discussions. Our analysis shows that the social support of accepted tweets is typically considerably stronger than for rejected tweets. Also, the set of accepted tweets seems to be very stable regarding changes to the social support of the tweets, so, even when considering support propagation, we mainly observe differences in such set when using the more permissive probability thresholds.

Given a Twitter discussion the output of the system is the biggest set of tweets of the discussion which can be globally accepted according to the skeptical approach based on the ideal semantics of a valued abstract argumentation framework.

The ideal semantics for valued argumentation guarantees that the set of tweets in the solution is the maximal set of tweets that satisfies that it is consistent, in the sense that there are no defeaters among them, and that all of the tweets outside the solution are defeated by a tweet within the solution. That is, if a tweet outside the solution defeats a tweet within the solution, it is, in turn, defeated by another tweet within the solution. In other words, the solution is the biggest consistent set of tweets that defeats any defeaters outside the solution.

The defeat relationship between tweets is evaluated by combining the criticism and support relationships, according to a given uncertainty threshold, and taking into account the weight of tweets considering different information sources from the social network, such as the number of followers of the author, the number of retweets and the number of favorites. The system can be of special relevance for assessing Twitter discussions in fields where identifying groups of tweets globally compatible or consistent, but at the same time that are widely accepted, is of particular interest, such as for instance for the assistance and guidance of marketing and policy makers.

The rest of the paper is organized as follows. In the next subsection we summarize the more relevant related work within the framework of argumentation models for social context. In Section 2, we define the formal structure to model Twitter discussions, assigning probability values to the relationships between tweets expressing the degree of belief in them, and weights to the tweets expressing their social relevance. Then, in Section 3, we extend the reasoning system with the information provided by support relationships between tweets. Finally, in Section 4 we analyze some Twitter discussions and, in Section 5, we conclude.

1.1. Related work

The idea of considering the relevance of the arguments in argumentation systems applied to social networks has also been studied by Leite and Martins [32]. In their work, the authors propose a semantical extension of Dung's Abstract Argumentation Framework [14] called Social Abstract Argumentation Framework. This framework incorporates social voting by adding votes for and against arguments, where votes are assumed to be extracted from an online debating system and represent the arguments' strength. Later on, Egilmez et al. [18] extended the framework to incorporate voting on attacks, including a social notion of attack strengths. The semantics of Social Abstract Argumentation assigns one or more models to debates and is parameterized by a set of operators that characterize how votes should be interpreted, the effect of attacks, and how multiple attacks should be combined. In [13], the authors propose an iterative algorithm to approximate the models of debates structured according to Social Abstract Argumentation.

The exploitation of Twitter by means of argumentation frameworks has also been explored by Grosse et al. [24,25], who defined a framework which allows opinion mining from incrementally generated Twitter queries, triggering the construction of argument trees such as those found in classical Dialogue-based Argumentation [6]. In their approach, an argument is a set of tweets for a given query (mainly a set of hashtags), and a tree is a hierarchical relation between them, with subsumption and conflict relations. The trees obtained resemble dialectical trees used in their previous work on Defeasible Logic Programming [38], although no argumentation algorithm is defined to extract the most relevant arguments from trees.

Our system is close to the argumentation framework developed by Cabrio and Villata [8], where natural language debates are analyzed and the relations among the arguments are automatically extracted. The authors use Bipolar Argumentation algorithms and semantics to evaluate the set of accepted arguments, given the support and the attack relations among them. The arguments and the relations among them are detected by an automated framework by applying natural language techniques, since the system is focused on online debate such as Debatepedia. One key difference between our system and the one proposed by Cabrio and Villata is that we incorporate both weighted arguments and probabilistic valued relationships. Weights are computed from different attributes of a tweet, such as the number of followers of the author, the number of retweets and the number of favorites, while the probability values are computed from informal descriptions expressed in natural language, by means of an automatic labeling system based on Support Vector Machines. We believe that the incorporation of weights and degrees of belief to obtain the relative relevance of arguments and the belief in the attacks, respectively, considering information taken from the social network, is an important aspect if we eventually want to build tools that are useful for analyzing discussions, considering different sources of information for socially accepted arguments. Despite the fact that our argumentation system can be utilized to analyze discussions in different social networks, in this work we focused on the analysis of Twitter discussions that are

Please cite this article as: T. Alsinet et al., An argumentative approach for discovering relevant opinions in Twitter with probabilistic valued relationships, Pattern Recognition Letters (2017), http://dx.doi.org/10.1016/j.patrec.2017.07.004