



Extraction and clustering of arguing expressions in contentious text



Amine Trabelsi*, Osmar R. Zaiane

Department of Computing Science, University of Alberta, Edmonton T6G 2E8, Canada

ARTICLE INFO

Article history:

Received 9 January 2015

Accepted 28 May 2015

Available online 10 June 2015

Keywords:

Contention analysis

Topic models

Arguing expression detection

Opinion mining

Unsupervised clustering

Online debates

ABSTRACT

This work proposes an unsupervised method intended to enhance the quality of opinion mining in contentious text. It presents a Joint Topic Viewpoint (JTV) probabilistic model to analyze the underlying divergent arguing expressions that may be present in a collection of contentious documents. The conceived JTV has the potential of automatically carrying the tasks of extracting associated terms denoting an arguing expression, according to the hidden topics it discusses and the embedded viewpoint it voices. Furthermore, JTV's structure enables the unsupervised grouping of obtained arguing expressions according to their viewpoints, using a proposed constrained clustering algorithm which is an adapted version of the constrained k-means clustering (COP-KMEANS). Experiments are conducted on three types of contentious documents (polls, online debates and editorials), through six different contentious data sets. Quantitative evaluations of the topic modeling output, as well as the constrained clustering results show the effectiveness of the proposed method to fit the data and generate distinctive patterns of arguing expressions. Moreover, it empirically demonstrates a better clustering of arguing expressions over state-of-the-art and baseline methods. The qualitative analysis highlights the coherence of clustered arguing expressions of the same viewpoint and the divergence of opposing ones.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Sentiment analysis, also referred to as opinion mining, is an active research area in natural language processing as well as data mining, that aims to extract and examine opinions, attitudes and emotions expressed in text, with respect to some topic in blog posts, comments and reviews. In addition to sentiment expressed towards products, other online text sources such as opinion polls, debate websites and editorials may contain valuable opinion information articulated around some topics of contention. In this paper, we address the issue of improving the quality of opinion mining from contentious texts, found in surveys' responses, debate websites and editorials. Mining and summarizing these resources are crucial, especially when the opinion is related to a subject that stimulates divergent viewpoints within people (e.g., Healthcare Reform, Same-sex Marriage, Israel/Palestine conflict). We refer to such subjects as issues of contention. A *contentious issue* is “*likely to cause disagreement between people*” (cf. Oxford Dictionaries).¹ Documents such as survey reports, debate site posts and editorials may contain multiple contrastive viewpoints regarding a particular issue of contention. Table 1 presents an example of short-text documents expressing divergent opinions where each is exclusively supporting or opposing a healthcare legislation.²

* Corresponding author.

E-mail addresses: atrabels@ualberta.ca (A. Trabelsi), zaiane@ualberta.ca (O.R. Zaiane).

¹ <http://www.oxfordlearnersdictionaries.com/definition/english/contentious>.

² Extracted from a Gallup Inc. survey <http://www.gallup.com/poll/126521/favor-oppose-obama-healthcare-plan.aspx>.

Table 1
Excerpts of support and opposition opinion to a healthcare bill in the USA.

Support viewpoint
Many people do not have health care. Provide health care for 30 million people. The government should help old people.
Oppose viewpoint
The government should not be involved. It will produce too much debt. The bill would not help the people.

Opinion in contentious issues is often expressed implicitly, not necessarily through the usage of usual negative or positive opinion words, like “bad” or “great”. This makes its extraction a challenging task. It is usually conveyed through the arguing expression justifying the endorsement of a particular point of view. The act of arguing is “to give reasons why you think that something is right/wrong, true/not true, etc., especially to persuade people that you are right” (cf. Oxford Dictionaries). For example, the arguing expression “many people do not have healthcare”, in Table 1, implicitly explains that the reform is intended to fix the problem of uninsured people, and thus, the opinion is probably on the supporting side. On the other hand, the arguing expression “it will produce too much debt” denotes the negative consequence that may result from passing the bill, making it on the opposing side.

The automatic identification and clustering of these kinds of arguing expressions, according to the topics they invoke and the viewpoints they convey, are enticing for a variety of application domains. For instance, it can save journalists a substantial amount of work and provide them with drafting elements (viewpoints and associated arguing expressions) about controversial issues. Moreover, a good automatic browsing of divergent arguing expressions in a conflict/issue would help inquisitive people understand the issue itself (e.g., same-sex marriage). Also, it may be used by politicians to monitor the change in argumentation trends, i.e., changes in the main reasons expressed to oppose or support viewpoints. The significant changes may indicate the occurrence of an important event (e.g., a success of a politician’s action or speech). Automatic summarization of arguing expressions may benefit survey companies who usually collect large verbatim reports about people’s opinion regarding an issue of contention. From a text mining perspective, representing a contentious document, as a small set of dimensions, each corresponding to an arguing expression of a different topic and viewpoint, is useful for information retrieval tasks like query answering or dimensionality reduction. In addition, it would enhance the output quality of the opinion summarization task in general.

The rest of this paper is organized as follows. Section 2 states the problem. Section 3 explains the key issues in the context of recent related work. Section 4 provides the technical details of the proposed model, the Joint Topic Viewpoint model (JTV). Section 5 describes the clustering algorithm applied on JTV output and used to obtain a feasible solution. Section 6 provides a description of the experimental setup on three different types of contentious text. Section 7 quantitatively assesses the JTV and constrained clustering adequacy and compares the performance of our solution with state-of-the-art and baseline methods. Section 8 qualitatively evaluates the coherence and the relevance of the final output of our method. Section 9 discusses the future work.

2. Problem statement

This paper examines the task of mining the underlying topics and the hidden viewpoints of arguing expressions towards the summarization of contentious text. An example of a human-made summary of arguing expressions [1] on, what is commonly known as the Obama healthcare reform, is presented in Table 2. The ultimate research’s target is to automatically generate similar snippet-based summaries given a corpus of contentious documents. However, this paper tackles the initial sub-problem of identifying recurrent words and phrases expressing arguing and clusters them according to their topics and viewpoints. This would help solve the general problem. Indeed, the clustered words and phrases can be used as input to query the original documents via information retrieval methods in order to extract relevant fragments or snippets of text related to a particular arguing expression. We use

Table 2
Human-made summary of arguing expressions supporting and opposing ObamaCare.

Support viewpoint	Oppose viewpoint
People need health insurance/many uninsured System is broken/needs to be fixed. Costs are out of control/help control costs. Moral responsibility to provide/fair Would make healthcare more affordable Don’t trust insurance companies.	Will raise cost of insurance/less affordable Does not address real problems Need more information on how it works Against big government involvement (general) Government should not be involved in healthcare. Cost the government too much

Download English Version:

<https://daneshyari.com/en/article/378696>

Download Persian Version:

<https://daneshyari.com/article/378696>

[Daneshyari.com](https://daneshyari.com)