

Dimensions of Argumentation in Social Media

Jodi Schneider¹, Brian Davis¹, and Adam Wyner²

¹ Digital Enterprise Research Institute, National University of Ireland, Galway,
firstname.lastname@deri.org

² Department of Computer Science, University of Liverpool
A.Z.Wyner@liverpool.ac.uk

Abstract. Mining social media for opinions is important to governments and businesses. Current approaches focus on sentiment and opinion detection. Yet, people also justify their views, giving arguments. Understanding arguments in social media would yield richer knowledge about the views of individuals and collectives. Extracting arguments from social media is difficult. Messages appear to lack indicators for argument, document structure, or inter-document relationships. In social media, lexical variety, alternative spellings, multiple languages, and alternative punctuation are common. Social media also encompasses numerous genres. These aspects can confound the extraction of well-formed knowledge bases of argument. We chart out the various aspects in order to isolate them for further analysis and processing.

1 Introduction

In social media, people continually express their opinions. These opinions are used to help businesses understand their customers and for governments to understand citizen needs and desires. 80% of data on the Web and on internal corporate intranets is unstructured, hence analysing and structuring the data is a large and growing endeavour³. In our view, an important way in which the data can be analysed and further structured is in terms of argumentation. However, we first have to understand the dimensions of expression of argument, which can then be isolated for further analysis and processing. Besides driving the input to knowledge bases, argumentation can also be used for the output of knowledge bases, providing justification and explanation.

Consistency in knowledge bases is essential since we cannot draw informative inferences with inconsistent knowledge bases. In social media, it is obvious that there is lots of disputed information concerning matters of fact (what is or is not true) and of opinion (what individuals believe or prefer). To make use of the knowledge in social media and reason with it, we must treat inconsistency. While a knowledge base may be filtered or truncated based on heuristics, some inconsistencies may remain, whether explicitly or implicitly. Alternatively, users may resolve inconsistencies based on limited weighting information such as provenance or a preference ranking. But to decide which fact is correct or which opinion is most relevant to them, consumers need to go beyond such rankings and to understand how statements are justified and the sources of disagreement. For this, we believe argumentation theory is crucial.

³ <http://www.gartner.com/it/page.jsp?id=1454221>

Current approaches to extracting and retrieving information from social media use opinion summarisation (e.g. summing votes for or against), topic-based [8] and feature-based text summarisation [7], and visualisation [4]. Such approaches discover trends, relationships, and correlations in data. While they may record inconsistency, they do not provide the means to articulate an elaborate structure of justification and disagreement.

While social media records arguments, current information extraction and knowledge acquisition systems do not represent these arguments, hence people must assimilate and use them unaided. One approach in the direction of representing argument is stance detection [9], which concerns identifying which side a party is taking in a debate, and which responses are rebuttals. While this adds further substance, it does not enable identifying the structure and layers of rationales for and against a position.

Even though current approaches are highly useful in decision making, the whole chain of rationale may be crucial. The overall popularity of an opinion is not as important as the reasons supporting it: overwhelming numbers of people buying a product may not matter as much as a particular reason for *not* buying it. The issue is whether it is the right product for the buyer, which is a matter not only of the pros and cons, but also of the explanations and counterarguments given. In our view, current approaches detect problems, but obscure the chains of reasoning about them.

The challenge is to extract the arguments from the text, turning textual sources into a representation that we can reason with even in the face of inconsistency. We explore these issues as follows. In Section 2, we first introduce the goals of argumentation extraction and provide a sample problem. In Section 3, we outline formalisations of argumentation that enable reasoning with inconsistent data. However, we note the gap between the formalisation and the argument analysis and extraction from source material. This highlights the need for greater understanding of the dimensions of argumentation in the social media landscape, which we discuss in Section 4. In closing, we outline the next steps to bridge between textual sources and the target formal analysis.

2 Goals and Example

Our goal is to extract and reconstruct argumentation into formal representations which can be entered into a knowledge base. Drawing from existing approaches to subjectivity, topic identification, and knowledge extraction, we need to indicate disagreements and other relationships between opinions, along with justifications for opinions. This is currently done by hand. The goal really is to figure out how to automate the analysis. Issues include the informality of language in social media, the amount of implicit information, and various ‘meta’ information that contributes to the argument reconstruction, as we later discuss.

Consider the situation where a consumer wants to buy a camera. In reviews, there may be a high degree of negative sentiment related to the battery, which a consumer can use to decide whether or not she wants to buy the camera. Yet, in the comments to a discussion, we may find statements about whether or not this is in fact true, whether it outbalances other features of the camera, whether the problem can be overcome, and so on. It is not enough to say “you shouldn’t buy this camera” – one needs to give the reasons why. Then the debate becomes an argument about the justifications: “it’s lightweight, you should buy it”, “the lens sucks, you shouldn’t buy it”, “the lens

doesn't matter, it has a bad battery" and so on. The argument is not just point and counterpoint; it is also about how each premise is itself supported and attacked. Each of these justifications may be further discussed, until the discussion 'grounds out' with no further messages. This has the structure of an argument, where points and counterpoints are presented, each implied by premises, which themselves can be argued about further.

Thus we envision deepening the knowledge bases constructed from social media based on the justifications given for statements. To do so, we need to better understand how disagreements and justifications—which we refer to collectively as argumentation—are expressed in social media. However, we first consider our target formalisation.

3 Formalising Argumentation and Argumentation Schemes

Abstract argumentation frameworks have been well-developed to support reasoning with inconsistent information starting with [6] and much subsequent research ([1], [2], [3]). An abstract argument framework, as introduced by Dung, [6] is a pair $AF = \langle \mathcal{A}, attack \rangle$, where \mathcal{A} is a set of arguments and *attack* a binary relation on \mathcal{A} . A variety of semantics are available to evaluate the arguments. For example, where $AF = \langle \{A1, A2, A3, A6, A7\}, \{att(A6, A1), att(A1, A6), att(A7, A2)\} \rangle$, then the preferred extensions are: $\{A3, A6, A7\}$ and $\{A2, A3, A7\}$.

However, Dung's arguments are entirely abstract and the attack relation is stipulated. In other words, it is unclear why one argument attacks another argument, as there is no content to the arguments. In order to instantiate arguments we need argumentation schemes, which are presumptive patterns of reasoning [10].

An instantiated argumentation scheme, such as Position To Know, has a textual form such as: 1. Ms. Peters is in a position to know whether Mr. Jones was at the party. 2. Ms. Peters asserts that Mr. Jones was at the party. 3. Therefore, presumptively, Mr. Jones was at the party. This has a formal representation in a typed logical language with functions from argument objects to predicates. The language formally represents the propositions required of the scheme as well as aspects of defeasible reasoning [12].

While this is an attractive approach to tying textual arguments to abstract argumentation, it relies on abstracting away the context and auxiliary aspects. It is far from clear how an argument such as represented in Section 2 can be transformed into a formal argumentation scheme so that it can be reasoned in an argumentation framework. To make use of the formal analyses and related implemented tools for social media discussions, a range of additional issues must be considered, as we next discuss.

4 Dimensions of Expression

To extract well-formed knowledge bases of argument, we must first chart out the various dimensions of social media, to point the way towards the aspects that argumentation reconstruction will need to consider, so that we later can isolate these aspects.

Social media encompasses numerous **genres**, each with their own conversational styles, which affect what sort of rhetoric and arguments may be made. One key feature is the extent to which a medium is used for broadcasts (e.g. monologues) versus conversations (e.g. dialogues), and in each genre, a prototypical message or messages could be described, but these vary across genres due to social conventions and technical constraints. De Moor and Efimova compared rhetorical and argumentative aspects

of listservs and blogs, identifying features such as the likelihood that messages receive responses, and whether spaces are owned communities or by a single individual, and the timeline for replies [5]. Important message characteristics include the typical and allowable message length (e.g. space limitations on microblogs) and whether messages may be continually refined by a group (such as in StackOverflow).

Metadata associated with a post (such as poster, timestamp, and subject line for listservs) and additional structure (such as pingbacks and links for blogs) can also be used for argumentation. For example, a user's most recent post is generally taken to identify their current view, while relationships between messages can indicate a shared topic, and may be associated with agreement or disagreement.

Users are different, and **properties of users** are factors that contribute not only to substance of the user's comment, but as well to how they react to the comments of others. These include demographic information such as the user's age, gender, location, education, and so on. In a specific domain, additional user expectations or constraints could also be added. Different users are persuaded by different kinds of information. Therefore, to solve peoples' problems, based on knowledge bases, when dealing with inconsistency, understanding the purposes and goals that people have would be useful.

Therefore, the **goals of a particular dialogue** also matter. These have been considered in argumentation theory: Walton & Krabbe have categorized dialogue types based on the initial situation, participant's goal, and the goal of the dialogue [11]. The types they distinguish are inquiry, discovery, information seeking, deliberation, persuasion, negotiation and eristic. These are abstractions—any single conversation moves through various dialogue types. For example, a deliberation may be paused in order to delve into information seeking, then resumed once the needed information has been obtained.

Higher level **context** would also be useful: different amounts of **certainty** are needed for different purposes. Some of that is inherent in a task: Reasoning about what kind of medical treatment to seek for a long-term illness, based on PatientsLikeMe, requires more certainty than deciding what to buy based on product reviews.

Informal language is very typically found in social media. Generic language processing issues, with misspellings and abbreviations, slang, language mixing emoticons, and unusual use of punctuation, must be resolved in order to enable text mining (and subsequently argumentation mining) on informal language. Indirect forms of speech, such as sarcasm, irony, and innuendo, are also common. A step-by-step approach, focusing first on what *can* be handled, is necessary.

Another aspect of the informality is that **much information is left implicit**. Therefore, inferring from context is essential. **Elliptical statements** require us to infer common world knowledge, and connecting to existing knowledge bases will be needed.

We apply **sentiment techniques** to provide candidates for argumentation mining and especially to identify **textual markers of subjectivity and objectivity**. The arguments that are made about or against purported facts have a different form from the arguments that are made about opinions. Arguments about objective statements provide the reasons for believing a purported fact or how certain it is. Subjective arguments might indicate, for instance, which users would benefit from a service or product (those similar to the poster). Another area where subjective arguments may appear is discussions of the trust and credibility about the people making the arguments.

5 Conclusions

There is intense interest in extracting information from social media, and particularly in the views people express, and how they express agreement and disagreement, and justify their views. This motivates us to translate existing approaches for text analysis and argumentation mining into techniques for identifying and structuring arguments from social media [13]. But these tools and resources must first be adapted for differences in social media. Understanding these differences is a critical first step, therefore, we have discussed the dimensions of argumentation in social media. Our purpose has been to make explicit the various challenges, so that we can move towards creating knowledge bases of argumentation. Next, the challenges identified should be transformed into requirements.

Acknowledgements

The first and second authors' work was supported by Science Foundation Ireland under Grant No. SFI/09/CE/I1380 (Líon2). The third author was supported by the FP7-ICT-2009-4 Programme, IMPACT Project, Grant Agreement Number 247228. The views expressed are those of the authors.

References

1. T. J. M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.
2. A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.
3. M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.
4. C. Chen, F. Ibekwe-Sanjuan, E. San Juan, and C. Weaver. Visual analysis of conflicting opinions. In *Proceedings of IEEE Symposium on Visual Analytics Science and Technology (VAST)*, 2006.
5. A. de Moor and L. Efimova. An argumentation analysis of weblog conversations. In *The 9th International Working Conference on the Language-Action Perspective on Communication Modelling (LAP 2004)*, Rutgers University, 2004.
6. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2):321–357, 1995.
7. Y. Lu, C. Zhai, and N. Sundaresan. Rated aspect summarization of short comments. In *Proceedings of the 18th International Conference on World Wide Web (WWW '09)*, 2009.
8. I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *Proceedings of the 17th International Conference on World Wide Web (WWW '08)*, 2008.
9. M. A. Walker, P. Anand, R. Abbott, J. E. F. Tree, C. Martell, and J. King. That's your evidence?: Classifying stance in online political debate. *Decision Support Sciences*, 2011.
10. D. Walton. *Argumentation Schemes for Presumptive Reasoning*. Erlbaum, N.J., 1996.
11. D. N. Walton. *Commitment in dialogue*. State University of New York Press, Albany, 1995.
12. A. Wyner, K. Atkinson, and T. Bench-Capon. A functional perspective on argumentation schemes. In *Proceedings of Argumentation in Multi-Agent Systems (ArgMAS 2012)*, 2012.
13. A. Wyner, J. Schneider, K. Atkinson, and T. Bench-Capon. Semi-automated argumentative analysis of online product reviews. In *Proceedings of the Fourth International Conference on Computational Models of Argument (COMMA '12)*, 2012.