From Argument Diagrams to Argumentation Mining in Texts: A Survey

Andreas Peldszus, Applied Computational Linguistics, EB Cognitive Science, University of Potsdam, Potsdam, Germany
Manfred Stede, Applied Computational Linguistics, EB Cognitive Science, University of Potsdam, Potsdam, Germany

ABSTRACT

In this paper, the authors consider argument mining as the task of building a formal representation for an argumentative piece of text. Their goal is to provide a critical survey of the literature on both the resulting representations (i.e., argument diagramming techniques) and on the various aspects of the automatic analysis process. For representation, the authors also provide a synthesized proposal of a scheme that combines advantages from several of the earlier approaches; in addition, the authors discuss the relationship between representing argument structure and the rhetorical structure of texts in the sense of Mann and Thompsons (1988) RST. Then, for the argument mining problem, the authors also cover the literature on closely-related tasks that have been tackled in Computational Linguistics, because they think that these can contribute to more powerful argument mining systems than the first prototypes that were built in recent years. The paper concludes with the authors’ suggestions for the major challenges that should be addressed in the field of argument mining.

Keywords: Annotation Scheme, Argument Diagram, Argument Mining, Argumentation, Rhetorical Structure Theory, Theory of Argumentation Structure

1. INTRODUCTION: ANALYZING ARGUMENTATIVE TEXT

One of the central aspects of human communication is argumentation: the process of conveying inclinations, attitudes or opinions, and trying to make the partner accept them - or even adopt them. Cognitive agents, when they team up to solve a complex task, are in a similar position when the “division of labor” is to be negotiated. For human beings, the medium for arguing is natural language, whereas software agents do so in some suitable formal language. Nonetheless, with the progress that text mining techniques and applications have achieved in recent years, argumentation becomes increasingly relevant also for automatic processes, and hence also for cognitive computing.

DOI: 10.4018/jcini.2013010101
Among the considerable body of research in computational argument, surprisingly little attention has so far been devoted to the issue of locating and analyzing argumentation in naturally-occurring text. A presumable reason is the fact that argumentation in “real” text is often not particularly crisp and clean – the argument proper is being infiltrated with the full range of problems of linguistic expression that humans have at their disposal. On the other hand, especially with the growing importance of social media communication, the amount of written argumentative discourse is rapidly growing, and interesting practical applications of finding arguments in text become visible on the horizon.

In correspondence with the popular notion of text mining (nowadays a cover term for many classification, search, and information extraction tasks) and the thriving research field of opinion mining (the detection of sentiment or opinions on products, people, organizations, issues) we thus see argument mining as the automatic discovery of an argumentative text portion, and the identification of the relevant components of the argument presented there. For the purposes of this paper, we concentrate on the second of these two steps. That is, we assume that an argumentative text (or an argumentative portion of a text) is already at hand, and the goal is to produce an analysis of the underlying structure of the argument that is being presented — in other words, a labeling of text portions with their argumentative roles and their relations to one another.

Why would this endeavor be of any practical relevance? One example is given by Palau and Moens (2011), who work with legal texts and aim at tracing the argumentation put forward by the parties involved, which may significantly enrich the retrieval capabilities of legal databases. Then there is the aforementioned area of opinion mining, for which the identification of arguments presents a natural extension: In addition to finding out whether internet users like or dislike a particular product (or any other entity), one might very well be interested in the reasons those users give for their opinions, inclinations, or decisions. Beyond the obvious commercial perspective on this task, there is a (potentially more exciting) prospect of perusing argument mining for public deliberation, e.g., as a tool for assessing public opinion on political questions.

In this paper, we provide a survey of research that is relevant to the goals we just formulated. Our contribution consists of two parts. The first begins with a critical discussion of various proposals for argument annotation schemes (diagramming techniques) that have been made in the argumentation community. This includes some diagramming techniques whose primary use was the illustration of the “essence” of argument, but which (at least in our opinion) can also be applied to the analysis of naturally occurring text. Our analysis of these approaches in Section 2.1 leads us to a proposal for a new schema that combines some advantages of different earlier schemes (Section 2.2). As an add-on, Section 3 discusses the relation between argument structure annotation and more general text structure annotation, in particular as exemplified by Rhetorical Structure Theory (Mann & Thompson, 1988), which has been suggested by some researchers to be suitable for argument representation as well. — The second part of the paper then surveys work on automatic argument mining, which includes results on subtasks from neighboring disciplines in text analysis (Section 4). Finally, Section 5 offers conclusions and our view of the major challenges that work on argument mining should address in the near future.

2. ANNOTATION SCHEMES FOR ARGUMENTATION

In this section, we develop an annotation scheme for argumentation that draws on different ideas from the literature and our practical experiences with analyzing texts in the Potsdam Commentary Corpus (Stede, 2004). Naturally, the literature on argumentation is vast, and here we consider only work that has a clear focus on proposals for formal notations. A very useful
Related Content

Towards the Synergy of Cognitive Informatics, Neural Informatics, Brain Informatics, and Cognitive Computing
[www.igi-global.com/chapter/towards-synergy-cognitive-informatics-neural/72279?camid=4v1a](www.igi-global.com/chapter/towards-synergy-cognitive-informatics-neural/72279?camid=4v1a)

The Philosophical Approach Towards Wisdom
[www.igi-global.com/chapter/philosophical-approach-towards-wisdom/51435?camid=4v1a](www.igi-global.com/chapter/philosophical-approach-towards-wisdom/51435?camid=4v1a)
A Commonsense Approach to Representing Spatial Knowledge Between Extended Objects
Tiansi Dong (2009). Novel Approaches in Cognitive Informatics and Natural Intelligence (pp. 142-156).
www.igi-global.com/chapter/commonsense-approach-representing-spatial-knowledge/27305?camid=4v1a

Robust Feature Vector Set Using Higher Order Autocorrelation Coefficients
www.igi-global.com/article/robust-feature-vector-set-using/49691?camid=4v1a