Claims on demand – an initial demonstration of a system for automatic detection and polarity identification of context dependent claims in massive corpora

Ehud Aharoni
IBM Haifa Research Lab, Haifa, Israel

Carlos Alzate
IBM Dublin Research Lab, Ireland

Roy Bar-Haim
IBM Haifa Research Lab, Haifa, Israel

Yonatan Bilu
IBM Haifa Research Lab, Haifa, Israel

Lena Dankin
IBM Haifa Research Lab, Haifa, Israel

Iris Eiron
IBM Haifa Research Lab, Haifa, Israel

Daniel Hershcovich
IBM Haifa Research Lab, Haifa, Israel

Shay Hummel
IBM Haifa Research Lab, Haifa, Israel

Mitesh Khapra
IBM Bangalore Research Lab, India

Tamar Lavee[1]
IBM Haifa Research Lab, Haifa, Israel

Ran Levy
IBM Haifa Research Lab, Haifa, Israel

Paul Matchen
IBM YKT Research Lab, US

Anatoly Polnarov
Hebrew University, Jerusalem, Israel

Vikas Raykar
IBM Bangalore Research Lab, India

Ruty Rinott
IBM Haifa Research Lab, Haifa, Israel

Amrita Saha
IBM Bangalore Research Lab, India

Naama Zwerdling
IBM Haifa Research Lab, Haifa, Israel

David Konopnicki
IBM Haifa Research Lab, Haifa, Israel

Dan Gutfreund
IBM Haifa Research Lab, Haifa, Israel

Noam Slonim[2]
IBM Haifa Research Lab, Haifa, Israel

Abstract

While discussing a concrete controversial topic, most humans will find it challenging to swiftly raise a diverse set of convincing and relevant claims that should set the basis of their arguments. Here, we demonstrate the initial capabilities of a system that, given a controversial topic, can automatically pinpoint relevant claims in Wikipedia, determine their polarity with respect to the given topic, and articulate them per the user's request.

1 Introduction

The ability to argue in a persuasive manner is an important aspect of human interaction that naturally arises in various domains such as politics, marketing, law, and health-care. Furthermore, good decision making relies on the quality of the arguments being presented and the process by which they are resolved. Thus, it is not surprising that argumentation has long been a topic of interest in academic research, and different models have been proposed to capture the notion of an argument (Freeley and Steinberg, 2008).

A fundamental component which is common to all these models is the concept of *claim* (or *conclusion*). Specifically, at the heart of every argument lies a single claim, which is the assertion the argument aims to prove. Given a concrete topic, or context, most humans will find it challenging to swiftly raise a diverse set of convincing and relevant claims that should set the basis of their arguments.

[1] Present affiliation: Yahoo!

[2] Corresponding author, at noams@il.ibm.com

Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations, pages 6–9, Dublin, Ireland, August 23-29 2014.

In this work we demonstrate the initial capabilities of a system that, given a controversial topic, can automatically pinpoint relevant claims in Wikipedia, determine their polarity with respect to the given topic, and articulate them per the user's request.

2 Basic concepts and associated challenges

We define and rely on the following two concepts:
Topic: Short, usually controversial statement that defines the subject of interest.
Context Dependent Claim (CDC): General, and concise statement, that directly supports or contests the given Topic.

Given these definitions, as well as a few more detailed criteria to reduce the variability in the manually labeled data, human labelers were asked to detect CDCs for a diverse set of Topics, in relevant Wikipedia articles. The collected data were used to train and assess the performance of the statistical models that underlie our system. These data are freely available for academic research (Aharoni et al 2014).

The distinction between a CDC and other related texts can be quite subtle, as illustrated in Table 1. For example, automatically distinguishing a CDC like S1 from a statement that simply defines a relevant concept like S4, from a claim which is not relevant enough to the given Topic like S5, from a statement like S6 that merely repeats the given Topic in different words, or from a statement that represents a relevant claim which is not general enough like S7, is clearly challenging. Further, CDCs can be of different flavors, ranging from factual assertions like S1 to statements that are more of a matter of opinion (Pang and Lee 2008) like S2, adding to the complexity of the task. Moreover, our data suggest that even if one focuses on Wikipedia articles that are highly relevant to the given Topic, only ≈2% of their sentences include CDCs.

Furthermore, since CDCs are by definition concise statements, they typically do not span entire Wikipedia sentences but rather sub-sentences. This is illustrated in Table 2. There are many optional boundaries to consider when trying to identify the exact boundaries of a CDC within a typical Wikipedia sentence. This task further complicates the CDC detection problem. Thus, we are faced with a large number of candidate CDCs, of which only a tiny fraction represents positive examples that might be quite reminiscent of some of the negative examples. Finally, automatically determining the correct Pro/Con polarity of a candidate CDC with respect to the Topic poses additional unique challenges. Nonetheless, by breaking the problem into a set of modular tangible problems and by employing various techniques - specifically designed to the problems at hand - we obtain promising results, demonstrated by the capabilities of our system.

Topic	The sale of violent video games to minors should be banned
(Pro) CDC	*S1: Violent video games can increase children's aggression*
(Pro) CDC	*S2: Video game publishers unethically train children in the use of weapons* Note, that a valid CDC is not necessarily factual.
(Con) CDC	*S3: Violent games affect children positively*
Invalid CDC 1	*S4: Video game addiction is excessive or compulsive use of computer and video games that interferes with daily life.* This statement defines a concept relevant to the Topic, not a relevant claim.
Invalid CDC 2	*S5: Violent TV shows just mirror the violence that goes on in the real world.* This statement is not relevant enough to the Topic.
Invalid CDC 3	*S6: Violent video games should not be sold to children.* This statement simply repeats the Topic, and thus is not considered a valid CDC.
Invalid CDC 4	*S7: "Doom" has been blamed for nationally covered school shooting*. This statement is not general enough to represent a CDC, as it focuses on a specific single video game.

Table 1. Examples of CDCs and invalid CDCs.

*Because violence in video games is interactive and not passive, critics such as Dave Grossman and Jack Thompson argue that **violence in games hardens children to unethical acts**, calling first-person shooter games ``murder simulators'', although no conclusive evidence has supported this belief.*

Table 2. A CDC is often only a small part of a single Wikipedia sentence - e.g., the part marked in bold in this example.

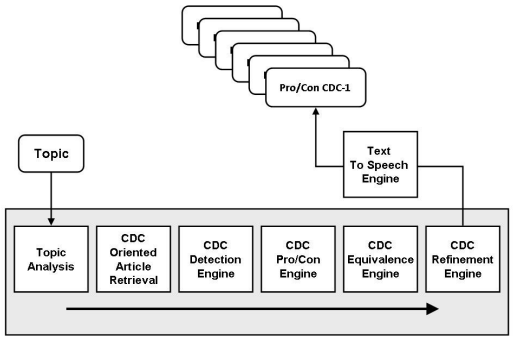

Figure 1. High level architecture of the demonstrated system.

3 High Level Architecture

The demonstrated system relies on a cascade of engines, depicted in Figure 1. In general, these engines rely on various IR, NLP and ML technologies, as well as different resources and lexicons like WordNet (Miller, 1995). Some engines are more mature than others, and, correspondingly, already employ a complex inner architecture, that will be discussed in more detail elsewhere. Given a Topic, the **Topic Analysis engine** starts with initial semantic analysis of the Topic, aiming to identify the main concepts mentioned in this Topic and the sentiment towards each concept. Next, the **CDC Oriented Article Retrieval** engine employs IR and opinion mining techniques in order to retrieve Wikipedia articles that with high probability contain CDCs. Next, the **CDC Detection engine** relies on a combination of NLP and ML techniques to zoom-in within the retrieved articles and detect candidate CDCs. A detailed description of this engine can be found in (Levy et al 2014). Next, the **CDC Pro/Con engine** aims to automatically determine the polarity of the candidate CDC with respect to the given Topic by analyzing and contrasting the sentiment towards key concepts mentioned in the Topic and within the candidate CDC. Next, the **CDC Equivalence engine** uses techniques reminiscent of automatic paraphrase detection to identify whether two candidate CDCs are semantically equivalent, so to avoid redundancy in the generated output. Finally, the **CDC Refinement engine** aims to improve the precision of the generated output, based on the results collected thus far; e.g., using a simple rule-based approach, we remove candidate CDCs for which the predicted Pro/Con polarity has low confidence. The remaining predictions are sent to the **Text To Speech engine** that articulates the top CDC predictions at the user's request.

4 Summary

Given a Topic, the demonstrated system is currently focused on detecting and articulating relevant CDCs. Combining this system with technologies that could automatically detect evidence to support these CDCs, may give rise to a new generation of automatic argumentation systems. In principle, such systems may swiftly detect relevant CDCs in massive corpora, and support these CDCs with evidence detected within other articles, or even within entirely different corpora, ending up with automatically

generated arguments that were never explicitly proposed before in this form by humans. The system described herein represents an important step in pursuing this vision.

Acknowledgments

We would like to thank our colleagues in the IBM debating technologies team who participate in generating more advanced versions of the system presented in this work, and further continue to contribute to essential aspects of this project. These include Priyanka Agrawal, Indrajit Bhattacharya, Feng Cao, Lea Deleris, Francesco Dinuzzo, Liat Ein-Dor, Ron Hoory, Hui Jia Zhu, Qiong Kai Xu, Abhishek Kumar, Ofer Lavi, Naftali Liberman, Yosi Mass, Yuan Ni, Asaf Rendel, Haggai Roitman, Bogdan Sacaleanu, Dafna Sheinwald, Eyal Shnarch, Mathieu Sinn, Orith Toledo-Ronen, Doron Veltzer and Yoav Katz.

References

Austin J. Freeley and David L. Steinberg. 2008. *Argumentation and Debate*. Wadsworth, Belmont, California.

Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. "A Benchmark Dataset for Automatic Detection of Claims and Evidence in the Context of Controversial Topics", in *Proceedings of the First Workshop on Argumentation and Computation, ACL 2014, to appear.*

Bo Pang and Lillian Lee. 2008. "Opinion mining and sentiment analysis" in *Foundations and Trends in Information Retrieval*, Vol. 2, pp. 1-135.

George A Miller. 1995. " WordNet: A Lexical Database for English" in *Communications of the ACM*, Vol. 38, pp. 29-41.

Ran Levy, Yonatan Bilu, Daniel Hershcovich, Ehud Aharoni and Noam Slonim. 2014. “Context Dependent Claim Detection.” In *Proceedings of the* 25th *International Conference on Computational Linguistics, COLING 2014, to appear.*