Πανεπιστήμιο Κύπρου

UNIVERSITY OF CYPRUS
DEPARTMENT of COMPUTER SCIENCE

ΕΠΛ 680
Cognitive Programming

Argument Mining
from Reviews & Debates

Orestis Makarounas 965787

Demos Ioannou 968030

Abstract

In this paper, we will talk about a new and rapidly growing area of natural language processing and data mining called argument mining. Argument mining is natural continuation and evolution of opinion mining – Sentiment Analysis, where in these two areas the main goal is to extract attitudes and opinions people hold about services, products, events and people from several reviews. Argument mining, on the other hand, focuses not only on **what** people think about a certain service but also on extracting the arguments that support that opinion (**why**). (Budzynska & Villata, 2017)

When this technology and area matures and grows in its final level, we will be able to automatically extract and detect arguments from vast amount of reviews and debates and be able to give answers on why people made certain decisions such as why did they vote this candidate for president.

Introduction

Every day more and more data are generated, but they become of limited usefulness if we can't process them efficiently. This unlimited growth of data is a big problem and it's called big data. A solution to this problem is data mining, which refers to the activity of going through big data sets to look for relevant and useful information without having to go through the whole set.

Argument Mining is a more specific area of the wider area of text mining that will allow us to detect the argumentation of a review or a debate and its structure. This implies the detection of all the arguments involved

in the argumentation process, as well of the argument components and the relationships between them. The goal of argumentation mining, an evolving research field in computational linguistics, is to design methods capable of analyzing people's argumentation.

Many researchers have understood the importance of argument mining and are getting involved to this area, as it's proved to be highly valuable.

Natural Language Processing (NLP)

Natural Language Processing is the area that explores how computers will be able to understand natural language. Its main goal is to make computers be able to answer questions about the contents of the text and draw inferences and important information from the text.

Data Mining

Data Mining is a process which can be used by big companies such as Amazon, eBay that can turn raw data into useful information. Such companies can learn more about their customers beliefs and thoughts about their services and then they can create better strategies to increase their sales.

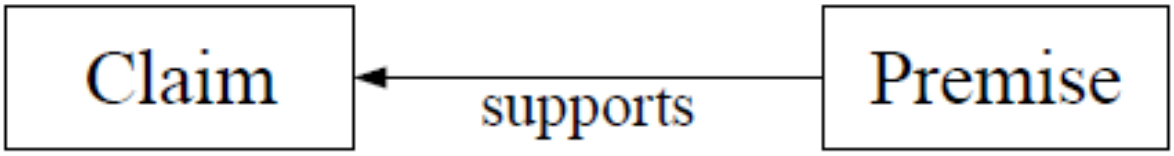
Argumentation

Argumentation has been studied since the early work of Aristotle, dating back to the 4th century BCE. It's a very important aspect in today's world and is included in many areas such as journalism, psychology, politics,

science, management, marketing etc. Most of the work of politicians, researches, managers and so on is based on identifying the pros and cons of a possible action and is based on the analysis of these arguments they take a decision.

Arguments

An argument is a reason or set of reasons given in support of an idea, action or theory. An argument form includes a claim or a conclusion, that is supported or attacked by premises. (Stab, Kirschner, Eckle-Kohler, & Gurevychyz)

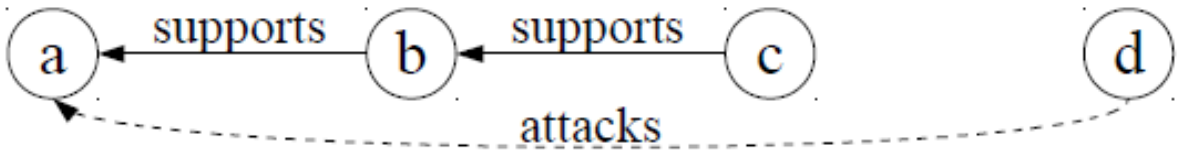


The claim which is the main component of the argument can be either true or false, so premises should support or attack the claim to prove the validity of the argument to the reader. In the following example we can see an argument, with the bold part of the argument being the claim, which is supported from the underlined part which is the premise.

"Everybody should study abroad (a).
It's an irreplaceable experience if you learn standing on your own feet (b) since you learn living without depending on anyone else (c). But one who is living overseas will of course struggle with loneliness, living away from family and friends (d):"

Argumentative relations are usually directed relations between two argument components and represent the **argumentation structure**.

The argumentation structure for our example is the following:



Premise (b) supports the claim (a) whereas premise (d) attacks the claim (a), also the premise (c) supports the premise (b). (Stab, Kirschner, Eckle-Kohler, & Gurevychyz)

Three important properties of argumentation structures:

1. Argumentative relations can hold between non-adjacent sentence/clauses, e.g. the argumentative attack relation from premise (d) to the claim (a).

2. Some argumentative relations are signaled by indicators, whereas others are not. For instance, the argumentative attack relation from premise (d) to the claim (a) is indicated by the discourse marker 'but', whereas the argumentative support relation from premise (b) to claim (a) is not indicated by a discourse marker.

3. Argumentative discourse might exhibit reasoning chains, e.g. the chain constituted between argument components a, b, and c.

ARGUMENT mining (also referred to or associated with argumentation mining, computational argumentation or debating technologies) is the automatic identification and extraction of argument data from large resources of natural language texts. (Budzynska & Villata, 2017)

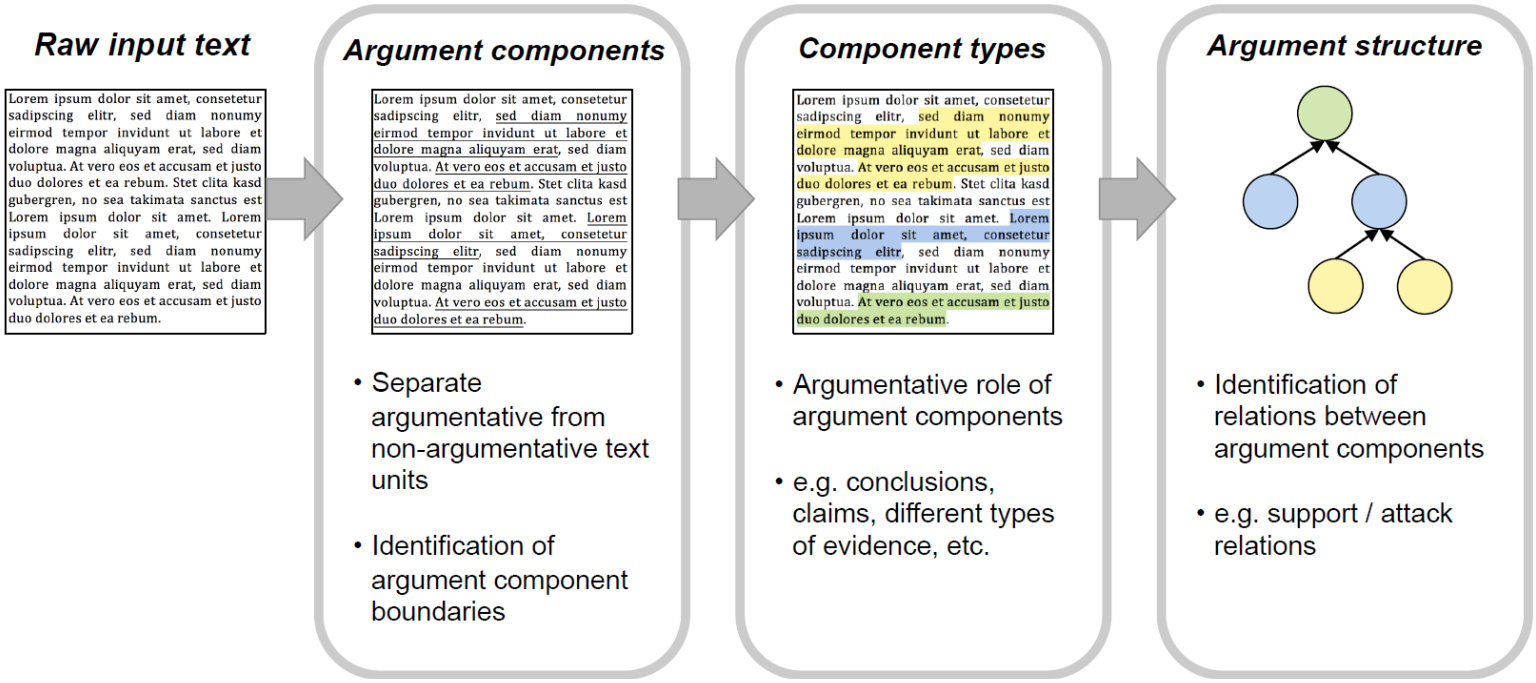
Argument mining is a natural continuation and evolution of sentiment analysis and opinion mining – two areas of text mining which became very successful and important both academically and commercially. In sentiment analysis, the work focuses on extracting people's attitudes (positive, neutral, negative) towards persons, events or products. In opinion mining, the work aims to mine people's opinions about persons, events or products. Argument mining, on the other hand, allows for recognizing not only what attitudes and opinions people hold, but also why they hold them.

The first step of an argumentation mining pipeline typically focuses on the identification of argumentative text units before analyzing the components or the structure of arguments. This task is usually considered as a binary classification task that labels a given text unit as argumentative or non-argumentative. Although the separation of argumentative from non-argumentative text units is an important step in argumentation mining, it merely enables the detection of text units relevant for argumentation and does not reveal the argumentative role of argument components.

The second step of an argumentation mining pipeline focuses on the classification of argument components aims at identifying the argumentative role (e.g. claims and premises) of argument components.

The third step of an argumentation mining pipeline focuses on understand the argumentative relations and build the argumentative structure. (Budzynska & Villata, 2017)

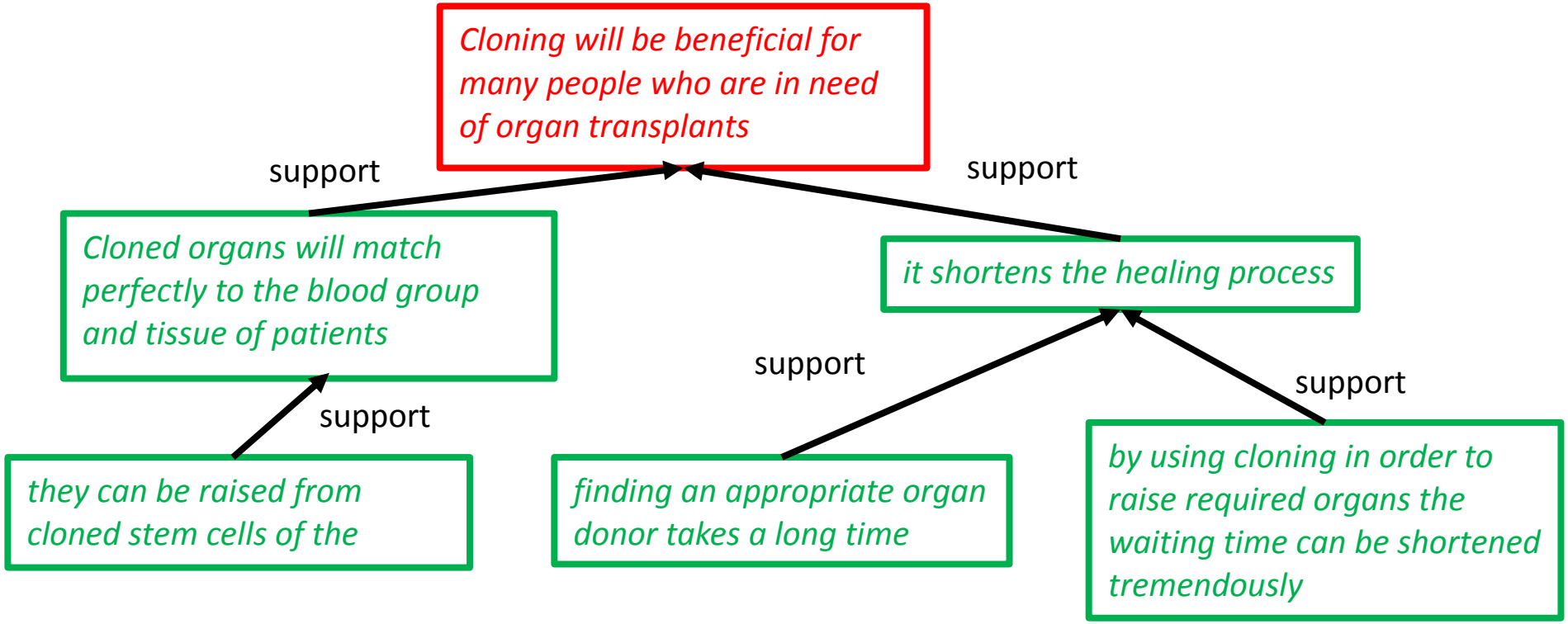
The following is a graphic representation of the above:



An example with text is the following: (Palau & Moens, 2009) (Stab & Gurevych, Annotating Argument Components and Relations in Persuasive Essays, 2014)

"Since researchers at the Roslin Institute in Edinburgh cloned an adult sheep, there is an ongoing debate if cloning technology is morally and ethically right or not. Cloning will be beneficial for many people who are in need of organ transplants. Cloned organs will match perfectly to the blood group and tissue of patients since they can be raised from cloned stem cells of the patient. In addition, it shortens the healing process. Usually, finding an appropriate organ donor takes a long time and by using cloning in order to raise required organs the waiting time can be shortened tremendously."

The underlined text is the argumentative text, where the red text is our claim and the green text is our premises. And the argumentative relations and argumentative structure is the below:



In this example we don't have attacks, if we had attacks we would have attack instead of support.

As we mentioned before, because of the large amount of data we have, we need an automated method to automatically identify argument components argumentative relation and to extract the arguments from a text. To do that we need supervised machine learning algorithms known as artificial neural networks. These algorithms need to be trained. The training data for our case are called corpora. The corpora are constructed manually by the annotators who annotate the input raw text. As the manual annotation is a highly time-consuming task, sharing and reusing analyzed data becomes a real value. (Budzynska & Villata, 2017)

The process of annotation starts with segmenting (splitting) the text into elementary discourse units (EDUs) or in fact into argumentative discourse units (ADUs). Annotators use software tools which help them to assign labels from the annotation scheme set to ADUs directly in a code. Next, the annotated data have to be stored as a corpus.

There are two basic styles of automation (in practice, they are often combined to form a hybrid approach): (a) the structural approach, i.e. grammars (hand coded set of rules) and (b) the statistical approach, i.e. machine learning (general learning algorithms).

In the structural approach, a linguist looks through a selected fragment of a corpus and aims to find patterns between different lexical cues in the text and categories in the annotation scheme. For instance, in a given corpus it might be observed that arguments are linguistically signaled by words such as “because”, “since”, “therefore”. Then, the linguist formulates rules describing these patterns in a grammar.

The statistical approach ‘replaces’ a linguist with an algorithm. In the same way as a human, a system will also look for patterns, however,

this time statistically on a larger sample of the training corpus which created manually by human annotators. The majority of the work in argument mining use the statistical approach. After training this algorithm is used to process raw, unannotated text in order to automatically extract arguments. These texts have to be the same as the set of texts which was manually annotated and stored as a test corpus. This step can be treated as an automated equivalent for manual annotation and corpus development.

Applications (Ubiquitous Knowledge Processing (UKP) Lab, 2018)

Some applications or projects for argument mining, especially in educational and research area are the Argumentative Writing Support. Current writing support is limited to feedback about grammar or spelling and there is no system that provides formative feedback about argumentative writing. We need to assist authors researchers or students in writing persuasive arguments with respect to the following questions:

- Is my argument well-structured and comprehensible?
- Are the given reasons relevant for my claim?
- Does my argument include sufficient support for being persuasive?

Argumentative Writing Support Goals are to:

- Create tools which aid in improving argumentation quality
- Develop methods for identifying argumentation structures in text
- Investigate novel models which automatically assess argumentation quality
- Provide formative feedback about argumentation

Another application is the Knowledge Extraction and Consolidation for Scientific Publications in the Educational Domain. Today, the number of publications increases heavily which leads to a large effort for researchers to find all relevant information on a topic. Therefore, we want to develop methods and tools which are able to automatically extract and consolidate knowledge from scientific publications in the exemplar domain of educational research.

The goals in this topic is to:

- Methods for extracting arguments from scientific publications will be developed. Arguments are a common instrument in scientific papers to give reasons for statements.
- We will explore methods to calculate relations between the extracted arguments. As a baseline we will use semantic similarity measures to find arguments which address a similar topic and to build up an argument graph containing all arguments and their relations as we mentioned previously. Based on this, we will explore textual entailment methods to get further relation types such as argument support or contradiction.
- Finally, it should be possible to search the extracted arguments for a given topic and to retrieve related arguments for a given argument.

Another application from Center for Argument Technology and BBC is to detect the fake news.(http://www.arg-tech.org/index.php/bbc-news-deploying-arg-tech-software/) BBC News are introducing School Report, an educational initiative for improving critical thinking skills and helping young people distinguish real news from fake. For 11 to 15-year-old, the BBC is partnering with Aardman Animations in providing an educational game around the theme of fake news. For the older age group of 16-18 year old, the BBC is partnering with the Centre for Argument Technology to deliver The Evidence Toolkit, for helping users analyze news articles to help them better understand what's fake and what's not. Let's take the first article in this site: https://evidence-toolkit-moral-maze.pilots.bbcconnectedstudio.co.uk/ , which talk about air pollution and if the diesel cars are dangerous. At the right of the page the reader takes instructions from the site. The first one is to "*Select the text that you think corresponds to the main claim of the article*". The second one is to "*Select each bit of text that you think corresponds to a reason (support) for the main claim*" and then for every one of these premises, as we called them before, it has to answer if it is "This is a piece of evidence. Is it presented as a fact or an opinion?" and if it is an opinion, is it comes from an expert on this topic? Because as we mentioned in our course, arguments from an expert are stronger due to expert's credibility. And it also asks if there is a counter-argument in this article. The reader, in order to answer the aforementioned questions, need to read this article very carefully. And this way of reading will make the new readers to read more carefully and thus to be able to detect the fake news.

Bibliography

Budzynska, K., & Villata, S. (2017). *Argument Mining.* IEEE Intelligent Informatics Bulletin.

Habernal, I., & Gurevych, I. (2016, June 14). Argumentation Mining in User-Generated Web Discourse. *Computational Linguistics*, pp. 125-179.

Palau, R. M., & Moens, M.-F. (2009). Argumentation Mining: The Detection, Classification and Structure of Arguments in Text. Barcelona, Spain.

Stab, C., & Gurevych, I. (2014). Annotating Argument Components and Relations in Persuasive Essays. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics*, (pp. 1501–1510). Dublin, Ireland.

Stab, C., & Gurevych, I. (2014). Identifying Argumentative Discourse Structures in Persuasive Essays. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 46–56). Doha, Qatar: Association for Computational Linguistics.

Stab, C., Kirschner, C., Eckle-Kohler, J., & Gurevychyz, I. (n.d.). Argumentation Mining in Persuasive Essays and Scientific Articles from the Discourse Structure Perspective. www.ukp.tu-darmstadt.de.

Ubiquitous Knowledge Processing (UKP) Lab. (2018, April 10). Retrieved from https://www.ukp.tu-darmstadt.de/ukp-home/research-areas/argumentation-mining/