Argumentation Mining: Annotating Argument Elements and Challenges Involved in Annotating Parliamentary Debates

Vaibhav D. Desai
HOD (BCA)
SDJ International College, Surat, India

Dr. Vishal Dahiya
HOD (MCA)
Indus University, Ahmedabad, India

Dr. Hiren Joshi
Professor
Computer Science Dept, Gujarat University, Ahmedabad, India

ABSTRACT

Argumentation mining focuses on automatic detection, classification and structuring the argumentation in text. Various state of the art machine learning techniques and context free grammars are applied to solve the challenges of argumentation mining. Research has been done in area of legal arguments, newspaper articles and social media blogs etc. This paper discusses main issues and challenges involved in identification of argument elements of parliamentary hansard. Identification of arguments is generally done by annotating argument components. The paper also discusses different techniques available for identification of argumentation elements and issues involved in it. It also highlights fundamental questions found during our research and enlist issues for future research.

KEYWORDS

Argumentation, parliamentary debates, argument elements, annotation

I. INTRODUCTION

Arguments are constructed and handled by the process called argumentation. Argumentation is extremely important for human being in

1) to understand new problems

2) to perform scientific reasoning

3) to express, to clarify and to defend their opinions in their daily lives.

This paper is organized as follows: Chap. 2 describes the fundamentals related to argumentation mining together with some formal definitions in this research area. Chap. 3 discusses various research made in this area as of today and future scope of research area. Chap. 4 discusses basics of annotation process & challenges involved in annotating parliamentary hansards. We conclude in Chap. 5 with discussion remarks and findings.

II ARGUMENTATION MINING

Argumentation mining is an area that covers different areas like; natural language processing, argumentation theory and information retrieval. The aim of argumentation mining is not to analyze the validity of an argument, but to automatically detect the argumentation of a document i.e. detection of all the arguments involved in the argumentation process, their individual or local structure and the relation between them[2]. This paper focuses on the detection of arguments from text data by annotating the components of an argument and the challenges involved in it.

Arguments are discussions in which reasons (Premises) are advanced for and against some proposition or proposal (conclusion)[1]. In other words, an argument is a communication presenting reasons for accepting a conclusion[3]. Arguments are different from proofs in a way that proofs are collected in a support of a claim which requires to be proven correct. Whereas arguments are non-monotonic and can be disproven[4].
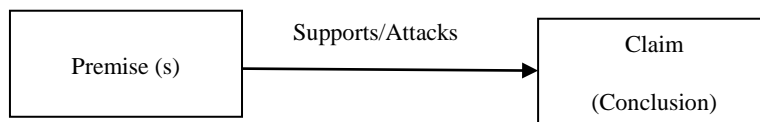
Figure 1: Simplest form of an argument

Some basic example patterns found in written arguments are as follows:

<claim> because <premise>

Since <premise> it is obvious that <claim>

<premise>. Therefore, <claim>

Even in some of the simple arguments, it is quite obvious that an argument is not limited to a single sentence. A single argument may contain multiple sentences.

III RELATED RESEARCH

In recent times, argumentation is tried to be automated, analyzed and performed by computers within the field of artificial intelligence. Argumentation has been a key area of importance in artificial intelligence and Law over last decade. Arguments found in judicial decision corpora are often complex and difficult to automate[5]. Argumentation research has been done in depth in newspaper, magazine articles, as the availability of the test data is relatively easy. In recent times automatic detection of argumentation is explored in area of social networking blogs, where people are fee to express their opinions. In some of the areas the rating (either negative or positive) is also possible based on the persuasiveness of detected arguments.

The detection of all the arguments presented in a free text is similar to the binary classification of all the propositions of the text as argumentative or non argumentative. There are two two types of problems associated with it: segmentation problem and classification problem. Segmentation problem is related with the boundary of argument i.e. where it starts and where it ends in the input data. Classification problem deals with a simple question, whether the sentence can be classified as argumentative statement or not. Hence, the first step in argumentation mining is the detection of argumentative sentences in the input document. This means that they contain an argument or a partial argument. This is generally implemented by binary classifiers, with an aim to discard the part of data which are not argumentative. The second step is to derive various argument components like claims, premises etc. At present various approaches like Naïve Bayse classifiers, Support Vector Machines, Maximum Entropy Classifiers, Decision Trees and Random Forests are used in machine learning.[5] Various approaches applied to same data result in contradictory result. This can be avoided by the proper choice of the features, based on which arguments are detected.

IV ANNOTATION PROCESS & CHALLENGES INVOLVED IN ANNOTATING PARLIAMENT HANSARDS

The annotation process is the process of identifying the part of data which are argumentative and tagging them with an appropriate tag. This process has mainly two steps: [6]

1. Annotating argument components i.e. (major-) claims and premises in each paragraph of data
2. These components are linked using attack and support relations i.e. which premise supports claims or attack claims

As we focus to identify the arguments from the input data, we shall focus mainly on step 1 of the annotation process.

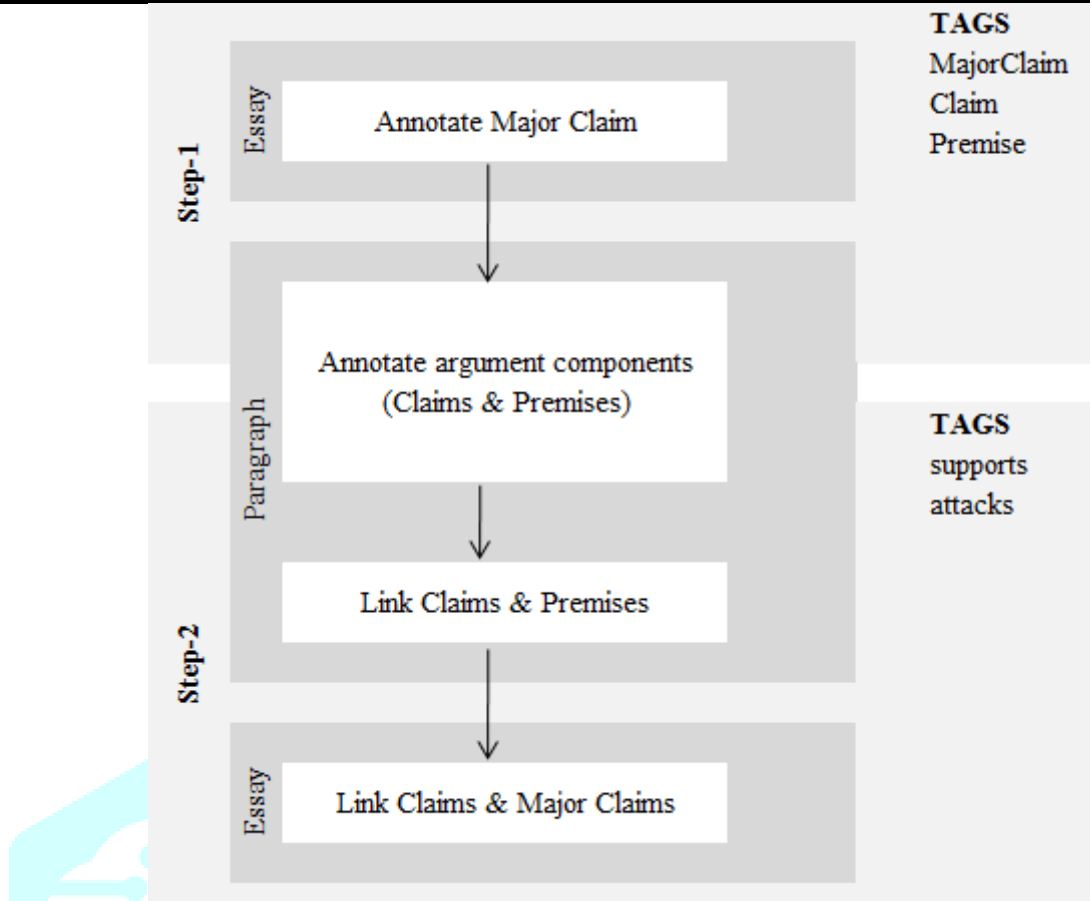Following figure illustrates the annotation process:

Figure 2: Overview of the annotation process

Step-1: Annotation of argument components

In this step, we focus on the clause level to identify the argument elements i.e. claims and premises. Although the author's stance and the context are very important for identifying these components, it is very likely to identify these components and tag them with an appropriate tag like major claim, claim and premise[6]. The following section describes the annotation for each tag in detail.

Step 1.1 Annotation of the Major Claim

Major claims are generally related to the opinionated expression with respect to the topic the author is taking about. It generally expresses the firm belief of the author and clarifies the stance author wants to take. For example, expressions like "I am against", "In my opinion" or "I strongly believe that" indicates the presence of a major claim in persuasive essays. One can found major claim in the introduction or in the conclusion of an input dataset. In the introduction, the major claim has the characteristics of a general assertion or an opinion with respect to the topic, whereas in the conclusion the major claim summarizes the argument according to the author's stance.[6]

There is a possibility that author does not include a major claim at all. This can generally happen when the author takes a neutral stance in an argument. So it is not mandatory that "MajorClaim" tag is available in each and every persuasive dataset. Major claims are generally identified at essay level.

One can prepare a list of major claim indicators which may vary based on the type of input data and stance the author of the data takes.

Step 1.2 Annotation of Claims
The central component of an argument is the claim or conclusion. [4] It is a claim that should not be accepted without support given by set of reasons called premises. Claims are generally identified at paragraph level. A claim can appear as conclusion

where its followed by a set of premises in the paragraph. Following are the indicators for the claims which are visible at conclusion level: "Therefore,". "For this reason", "In conclusion". A claim can also be found in the beginning of a paragraph, before any premises. This has the characteristics of initial assertion. For such claim indicator example would be "because", "for following reasons". A claim has to be tagged in either case that is tagged as assertion in the beginning of a paragraph or as a conclusion at the end of the paragraph. There are certain obvious claim indicators which are listed as below.

Accordingly	indicates that
as a result	I think must
consequently	I mean
Concludes	I believe
follows that	it should be clear that
hence	it is my contention
that clearly demonstrates	it is highly probable that
however	it follows that
in conclusion	be that
in short	on the contrary
in my opinion	proves that
in fact	so
Implies	Thus
and many more …….	

Table 1: Claim Indicators

Step 1.3 Annotation of premises

Premises are the reasons that supports or attacks the claim made in any argument. Both supporting and attacking premises are to be tagged with premise indicator. It is very much possible that a claim and premises appear in a single statement. Hence, premises are also annotated at the clause level. Just like claim indicators, there are set of premise indicators available for tagging. Following are the most common premise indicators:

after all	for example
assuming that	for instance
As	for one thing
as indicated by	for the reason that
as shown	furthermore given
besides	in light of
because	in view of
deduced	in view of the fact that
derived from	indicated by
due to	is supported by
firstly	secondly
follows from	since
for	whereas

Table 2: Premise Indicators

We can take an example from Canada parliament hansard for tagging. The following input dataset is from 42nd Canada parliament hansard dated 28th Jan 2016: [7]

CANADA ELECTIONS ACT

Mr. Don Davies (Vancouver Kingsway, NDP) moved for leave to introduce Bill, An Act to amend the Canada Elections Act (Voting age).

He said: Mr. Speaker, I am honoured to stand in the House to introduce a bill that would widen the franchise of this country by extending the privilege of voting to Canadians aged 16 or over, with great thanks to the member for Cowichan-Maklahat-Langford.

The history of franchise in this country is one of expansion. At one time only men could vote, only men with property. Women could not vote, first nations could not vote and people had to be 21 years of age. Studies show that individuals who begin voting early in our democratic process are more likely to continue voting for the rest of their lives. We know that voter turnout is generally anywhere between 50% and 65%. Therefore, this is an important initiative to get young voters engaged in our process.

- MajorClaim
  - To introduce a bill that would widen the franchise of this country
- Claim
  - This is an important initiative to get young voters engaged in our process
- Premise
  - Studies shows that individuals who begin voting early in our democratic process are more likely to continue voting for the rest of their lives.
  - We know that voter turnout is generally anywhere between 50% and 65%

Step 2: Annotation of Argumentative relations

Once the elements of arguments are identified and tagged appropriately, the next step would be establishing the relation between claims and premises. They can be tagged with tags like "supports" or "attacks". Relation between claims and major claims can also be done with similar tags. Once the annotation of argumentative relations has been done, the overall persuasiveness of an argument can be done. Since our research paper focuses on classification problem, this particular step is not discussed in detail.

CHALLENGES IN PARLIAMENTARY DEBATES

Parliamentary debate is an area where very little work has been done in argumentation mining. Unlike other areas we have discussed earlier, there are some additional challenges involved. Most of the countries at present are having multi-lingual debate hansard. India is a perfect example for this. Indian parliament is having a hansard which is written in 15+ languages. A single debate document is written in more than one language. The data available is not even semi-structured in most of the cases, making it too difficult for the researcher for mining it. Some parliaments are not even having text data of the debates. Instead, they are having audio data only which needs to be converted to textual data first and then it could be used for argumentation mining purpose. Some of the countries are having debates data soft copy after a certain time. The initial data is either not available or it is available in hard copies, where data may be partially available. All above limitations make it very difficult to annotate the debates as it needs extreme efforts in converting and cleansing the data so that it could be used for annotation.

V CONCLUSION

Argumentation mining is relatively new research area that does not focus on validity of the argument i.e. correctness of argument. Instead, it focuses mainly on automatic detection of argument elements and relation between the elements. An argument has

different structure then a non-argument statement(s). Annotation is the first step to identify the argument from persuasive input data. Different POS (Part Of Speech) indicators are available to annotate the data, which can vary based on the type of input data. Parliamentary debates are having its own challenges in automatic detection of argument elements.

REFERENCES

[1] Raquel Mochales and Marie-Francine Moens (April 2011) "Argumentation Mining" Springer Science & Business Media

[2] Katarzyna Budzynsca, Mathilde Janier, Juyeon Kang, Chris Reed, Patrick Saint-Dizier and Olena Yaskorska (May 2012) "Towards Argument Mining from Dialogue" School of Computing, University of Dundee, UK.

[3] Jodi Snchneider (Feb 2012) "Automated argumentation mining to the rescue? Envisioning argumentation and decision making support for debates in open online collaboration communities" INRIA Sophia Antipolis, France

[4] Raquel Mochales and Marie-Francine Moens (May 2009) " Argumentation Mining: The Detection, Classification and Structure of Arguments in Text" ICAIL-2009, Barcelona, Spain

[5] Macro Lippi and Paolo Torroni (April 2014) "Argument Mining: A machine Learning Perspective" ACL workshop ArgMining2014, SICSA

[6] Christian Stab and Iryna Gurevych (Nov 2018) "Annotating Argument Components and Relations in Pursuative Essays" Ubiquitos Knowledge Processing Lab, Germany

[7] http://www.ourcommons.ca/DocumentViewer/en/42-1/house/sitting-11/hansard