

Research on the Construction of Chinese Argument Corpus

Chen Zhang

Dalian Ocean University, Dalian, China

Email: 623347494@qq.com

How to cite this paper: Zhang, C. (2022). Research on the Construction of Chinese Argument Corpus. *Open Journal of Modern Linguistics*, 12, 123-138. <https://doi.org/10.4236/ojml.2022.121010>

Received: February 12, 2022

Accepted: February 22, 2022

Published: February 25, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Argument mining is an important task of natural language processing. As a branch of sentiment computing, its task is to automatically extract arguments from unstructured text documents in order to provide structured data for machine learning and deep learning models. It has recently become a hot topic because of its potential to process information from the Internet in an innovative way, especially its ability to process information from social media. Faced with the serious shortage of annotated corpora used to train supervised learning algorithms in the field of argumentation mining, we created a reliable annotated corpus of Chinese argument structures. In addition to the argument-oriented mining task, the close relationship between the direction of sentiment calculation and its consideration is also considered. At the same time, the data set can also be used as a sentiment classification task.

Keywords

Argumentation Mining, Corpus, Emotion Classification

1. Introduction

Argumentative digging, also known as argumentative analysis. When we express our position on a controversial topic and use argumentation to achieve the goal of persuasion, the process is called argumentation.

At the beginning of the twenty-first century, Computational Argumentation (CA) was born as an emerging research direction guided by research arguments in the field of Artificial Intelligence (AI). Since then, Computational Argumentation has had as its main goal the improvement of automatic reasoning ability of AI, trying to achieve the improvement by combining computation with human cognitive models about logical argumentation.

One of the important components in natural language processing is argument

mining, which is performed with the aim of automatically extracting structured arguments, also known as arguments, from unstructured text documents, allowing for easier application of structured data to computational models and arguments in inference engines (Xu, 2016). It recently becomes a hot topic because of its potential in handling information from the web in an innovative way, especially its potential for handling information from social media. It prefers to think of Argument as a textual structure, and the process of constructing an argument structure is represented by Argumentation, then the task of identifying and extracting arguments from textual data is called Argumentation Mining. A typical argument structure consists of a Premise, a Claim, and an inference rule, and then Argumentation can be referred to as the process in which the premise can effectively support the claim after using the corresponding inference rule, and Argumentation Mining is the operation of mining all the arguments and claims in the text data and then establishing logical connections between them.

The Web is today a very important source of data for most disciplines interested in this type of research, especially social media. Online newspapers, product reviews, blogs, etc. provide a heterogeneous and growing stream of information where (user-generated) arguments (arguments) can be discovered, extracted and analyzed. The availability of this data, coupled with tremendous advances in computational linguistics and machine learning, has created fertile ground for the emergence of a new field of research.

The research on argument mining is still in its infancy and still faces great difficulties and challenges in many aspects, such as: the lack of effective argument annotation models and standardized argument annotation methods; and the extremely scarce corpus, especially the high-quality and general annotated corpus (most of the corpus use standardized corpus, such as legal documents and argumentative essays, as annotation materials, which are small in size and not high in quality) restrict the development of argument mining techniques.

When reading the references, it is easy to find that most of the argumentation mining research done in China currently uses foreign argumentation-related corpus because it has a more complete annotation system and is recognized by most research institutions. However, the linguistic structure of English and Chinese is very different in some linguistic environments, which makes the Chinese corpus more difficult to annotate and define, and therefore leads to a lack of Chinese corpus.

It can be seen that foreign research and resource construction in the direction of argumentation mining can be broadly borrowed from each other, and gradually formed a distinctive annotation scheme and a more standardized annotation method. In contrast, the Chinese argumentative corpus in China is much scarcer, and there are different definitions of corpus selection and annotation schemes for different construction purposes. Therefore, in order to try to solve the above problems, the goal of this paper is to build a Chinese argumentative structure annotation corpus with reliable quality.

2. Related Theoretical and Technical Foundations

2.1. Argumentative Digging

Argumentation has always played an important role in human discourse, whether written or oral. To argue means to claim that something is true and reliable and to try to convince others that one's side of the argument is true and valid by providing arguments to support it. Argumentation is probably as old as mankind, because in everyday communication, people have to go through the process of arguing when they try to convince or inform others of certain findings. Throughout history, philosophers have studied argumentation. Argumentation was central to Western education from ancient Greece until the end of the 19th century. Orators and writers were trained to argue in order to persuade their audiences. The method of argumentation was based on rhetoric and logic, and by the 1950s it became a required part of university education.

With the advent of the Web 2.0 era and the rapid development of mobile Internet, users can publish and share their opinions on something or other on social networks, e-commerce sites, variety shows, official websites and other platforms. These opinions are often partly in support of the topic and partly expressing opposition, which contains a few neutral views. As debate is a common form of expression in daily life, it is important to find out the discursive body and its emotional tendency in the speaker's discourse in order to uncover the deeper information of the debate in the speaker's discourse. In this paper, the textual information comes from the speeches of the debaters in debate tournaments, which are more understandable than standard texts such as legal documents and argumentative essays, and contain a lot of argumentative structures. However, in the process of processing the text, the deeper speech of the debaters and the implied semantics in their discourse also make it difficult to mine the arguments for the debates.

Argumentation is defined as the act or process of constructing reasons and drawing conclusions and applying them to the case under discussion (Merriam-Webster). This act or process of providing reasons for or against something constitutes an important part of the argument. The arguments together with the conclusion or claim form the complete argument. Thus, the main components of an argument include.

- 1) Argument: also known as a claim or conclusion, i.e. something that people are for or against.
- 2) Argument: the evidence that people use to support the argument.
- 3) Rules of reasoning that link the initial claim to the argument and ensure that the workings of the argument can be understood.

However, the structure of arguments in reality is often more complex. Arguments may involve chains of reasoning in which an argument and its arguments are used as arguments to reason out a more general conclusion, thus forming a recursive tree structure. Arguments and arguments may also form other more complex graph structures, etc., which have been well studied in the argumenta-

tion literature.

In the age of big data, argumentative structures are found in all kinds of text and spoken language, such as legal texts and court decisions, medical cases, scientific articles, patents, reviews, online forums, user-generated content, debates, interactions, and conversations, to name a few. In today's information overload, there is a desire to build tools that help users quickly find arguments that support a statement or conclusion without having to read a lot of information. Argument mining, which includes the detection of argumentative structures in natural language text or speech, and the identification or classification of argumentative components, improves search and information retrieval tasks and aims to provide users with guided visualizations and summaries of argumentative structures in order to help them achieve easy and fast understanding.

2.2. Argument Structure

When a statement is presented, disagreements and debates arise about the content of the statement, and eventually the disagreements are removed by means of debate to reach some degree of consensus. This is how argumentation is used in philosophy and dialectics. Moreover, argumentation has been associated with various fields such as computer science, law, linguistics, and philosophy. The models of argumentation collected in the literature are rich, and although they are oriented to the English language, commonalities can still be found in them.

There are few large publicly available datasets for argument mining tasks. Araucaria DB is one of the most representative and comprehensive corpora in the field of argument mining (Reed et al., 2008), and it contains a wide variety of article types, such as newspaper editorials, parliamentary records, judicial briefs, and online discussions. The publicly available datasets released by Walker et al. in 2012 Resources such as the Internet Argument Corpus (IAC) and AIFdb released by Lawrence's team in 2012 provide us with domain-relevant datasets for high-quality argument analysis, but they do not provide the large amount of data and training of powerful classifiers needed for research, especially in the context of a particular topic or domain.

Toulmin's model (Toulmin, 1958) is one of the classic models of argument mining, which proposes that the logical structure of an argument consists of six factors: Claim, Grounds, Warrant, Modality, Backing, and Rebuttal. The argument refers to the core argument or conclusion of a statement; factual material can be understood as arguments, which support the argument through arguments, such as legal documents, rules and regulations, explicit rules of a particular field, and established common sense; warrants are used to prove that previously presented arguments and arguments can be justified, and to prove the feasibility of the proposed argument; modals are adjectives of degree that describe the strength of an argument. They are used to support the validity of the guarantee, and are intertwined with the guarantee, the factual basis, and the argument; exceptions often indicate special cases and contrary cases, which con-

tain content that constrains the validity of the argument.

Another well-known application of the argumentative model is the abstract argumentative model and the structured argumentative model. The former is based on Dung's work, and the abstract argument model presents the then relatively novel idea of treating argument elements as atomic entities and containing no internal structure. The definition of argument structure and annotation schemes becomes particularly important when the goal of the task is to extract arguments from natural language. Therefore, argument mining usually employs a structured argument model and sometimes embeds a dialogue model as well. As we have learned in the previous paper, although there are very classical structured argument models such as Toulmin and Freeman (Freeman, 1991) that provide valuable information in this direction, argument mining can be oriented to many domains, and the models need to be improved to different degrees for different domains and different problems. And considering that the implementation of theoretical models in practice often becomes intractable due to problems such as the difficulty of defining the boundaries of argumentative elements, researchers tend to simplify the models partly for specific tasks.

Walton (Walton, 2009) proposes a definition of a given intuitive argument structure, given the existence of the core element argumentative arguments in the argument structure, consisting of a set of premises and arguments, an argument and a logical reasoning of the conclusion mentioned in the previous section. In other literature readings, the definitions of claim, center, and conclusion have similarities, with minor modifications for different domains and tasks. The premise is also often referred to as Reason or Evidence. Jo et al. (2018) proposed a neural system architecture to simulate argumentative dialogue, modeling the interplay between the arguments of opinion holders and challengers, and predicting whether arguments are successfully changed. Durmus & Cardie (2019) constructed a debate dataset and modeled the roles to improve the performance of predicting the debate winner's task.

The application of argument mining techniques in natural language texts can be divided into several basic tasks based on the argument structure: identification of argument components, whose subtasks include argument detection and argument detection; prediction of argument relations or argument structure, which is newly proposed in recent years to accomplish the task of argument generation through adversarial generative networks; identification of relations within and between argument components, etc. The specific types of relations are usually related to the specific types of relations are usually task-related and can be simple correlations or subdivided into support or opposition relations, etc.

2.3. Task Composition

The complete argument mining system requires a series of intrinsically related subtasks that form a complete structure: first, unstructured text is input to the system, the part of the input text that contains argument information is ex-

tracted, and the boundaries of the argument components are obtained by comparing the text with argument information with the text without argument information. The structure of the argument is also predicted before the structured annotated text is obtained. In this paper, we introduce two classic argument mining subtasks: argument component detection and argument structure prediction.

The main goal of the first phase of the argument mining system is to detect argumentative components in the input documents, and the detected entities will be represented as nodes in the argument structure graph. Most existing algorithms usually divide this problem into two sub-problems to solve: extraction of argumentative statements and detection of argument boundaries. The former is at the sentence level, while the latter corresponds to different granularity and will be solved in different ways.

The first is the extraction of argumentative statements. The problem can also be understood as a classification task, and in principle any type of machine learning classifier can make an attempt for this classification task. In fact, there are several different options on how to transform it: training a binary classifier to distinguish between argumentative and non-argumentative sentences, leaving the task of identifying specific argumentative component types (e.g., claims or reasons) for the next stage; training a multi-classifier to distinguish all argumentative component types present in the adopted argumentative model, which assumes that a sentence can only include at most one argumentative component type; training a set of binary classifiers, each of which corresponds to one argument component type in the adopted argument model, so that a sentence can contain more than one argument component type.

The purpose of detecting argumentation boundaries is to determine the exact boundary of each argumentation component, also known as Argument Unit detection or Argumentative Discourse Unit detection. The task to be solved at this stage is to detect the location of the start and end of the argumentative component in each sentence that has been predicted to be argumentative. A typical solution is to consider the boundary detection problem as a sequence labeling problem or segmentation problem, and then solve it using a strategy similar to that used to deal with sequence labeling problems. Overall, it is difficult to achieve accurate argumentation boundary detection.

The final step of argument mining is called argument structure prediction, which requires identifying the associations between the argument components that have been identified in the previous step. This is a very challenging task because of the high level knowledge representation and reasoning involved. The output of this phase is a graph of detected argument components connected to each other, and the edges of the graph can indicate different relationships such as implication, support or conflict. This structured representation clearly expresses the internal structure and logical relationships of the text, thus enabling a deep understanding of the text, which is useful for tasks such as the determination of argument quality and the search of arguments.

2.4. Labeling Quality Assessment

In a corpus annotation task, the task publisher usually assigns the same corpus to multiple annotators for annotation in order to improve the quality of annotation. How to judge the quality of annotation results? Generally speaking, the higher the consistency of the annotated answers means the more reliable the annotated corpus is and the better the quality of the annotation is, while on the other hand, if the annotated answers are more divergent, the quality of the annotation is poor and not suitable for adoption. There is a specific measure of agreement here, namely inter-rater agreement algorithms, which measure the degree of agreement among several annotators on the same text. There are four commonly used agreement algorithms: percentage agreement (Burns, 2014), Fleiss Kappa (Fleiss, 1971), Krippendorff's α (Krippendorff, 2007) and Krippendorff's α_U (Krippendorff, 2004).

3. Corpus Construction

3.1. Concept Definition

In order to explore better argument components and structures, we consider that the debaters in the debate are composed of both sides, and therefore contain not only logical arguments, but also rich information about sentiment polarity and sentiment categories. Therefore, adding “position” tag, “sentiment polarity” tag and “sentiment category” tag can better help this paper to perform argument mining and sentiment classification tasks.

Table 1 shows the labeling scheme. The selection of emotion categories was based on the “Emotion Ontology Library” published by Dalian University of Technology.

Through extensive viewing of the debates, this paper identifies a final Chinese argumentative corpus annotation scheme containing.

- 1) Theme: Debate topic.
- 2) Claim: A debater's core argument in a paragraph, stating his or her position and summarizing his or her speech as a whole, e.g., “Our view is that this is a face-saving society”. Usually there is only one argument in a paragraph, but there are cases where there is more than one argument.
- 3) Premise: The debater adds a statement to the argument he or she is making. After stating his or her position, the debater will use a statement to explain the position he or she has stated. For example, “OK, hello, everyone, our side today is that this is a society of faces. Because let's start with the year of the dog, all animals reproduce, and what do they rely on? By choosing a mate. How to choose a mate, we have to choose a good gene ah, choose a good gene how to get it? Look at its appearance”.
- 4) Instance: The debaters give examples through common sense as well as their own personal experiences, and support their arguments by giving examples. For example, “In my original time I had a single eyelid, and my mother

Table 1. Labeling scheme and interpretation of nouns.

Name	Corresponding	Note
Theme	Theme	Debate questions (sentence level)
Claim (argument/focus)	Claim	Core arguments of the debaters (sentence level)
Premise (arguments/premises)	Premise	Arguments to prove the thesis (sentence level)
Instance	Instance	Examples to aid and support the argument (sentence level)
Modality (modal words)	Modality	Description of the strength of the argument (sentence level)
Stance (position)	Support	Support Position
	Against	Opposition position
	Neutral	Neutral position
	Positive	The emotional bias of the debaters
Emotion (emotional polarity)	Negative	The emotional bias of the debaters
	Neu	The emotional bias of the debaters
	Joy	Happy (PA)
	Good	Respect (PD)
Category (emotional category)	anger	Anger (NA)
	grief	Sadness (NB)
	fear	Panic (NI)
	evil	Doubt (NL)
	alarm	Strange (PC)
	Rhetorical question (RH)	
Figures (rhetorical device)	Irony (IR)	
	Puns (HM)	Rhetorical devices used in the course of argumentation
	Metaphor (MP)	
	Proportional (PS)	
	Exaggeration (EG)	
Debater	Debater	Debater's name
Votes (number of votes)	Votes	Number of votes on both sides per round (expressed in tuple)
Final vote	Fvote	Final number of votes on both sides (expressed in tuple)
Results	Results	Debate results

thought that double eyelids would look better, and now that there is this medical technology, I'll be honest, had a double eyelid cut, and that healed quickly”.

5) Modality: Words that describe the strength of the core argument, e.g., “probably”, “possibly”, “should”, “must”, “seems”. Example: “You can be whole as Fan Bingbing in one day, but whether you can read as Lin Huiyin in one day is something we don't know, so we think the face seems to be less important in this case”.

Next is affective polarity, the definition of affective categories.

1) Stance: the position of the debater, generally divided into the positive position (For) and the opposing position (Against), a few cases will appear the moderator stand neutral position, so add a neutral (Neutral) position.

2) Emotion: The debater's emotion is either Positive or Negative.

3) Category: Emotion category, refer to “Emotion Ontology Library” of Dalian University of Technology.

4) Figures: The rhetorical techniques used by the debaters in the course of their speeches, mainly rhetorical questions, metaphors, prose, irony and other common rhetorical techniques.

In order to have a clearer argument structure when labeling, labels with similar meaning or potential for confusion need to be more clearly defined. This thesis defines two types of argumentative relations between Claim and Premise: support and against, both of which are established by the following rule: an argument (Premise) must support or oppose an argument, but an argument often does not contain a position for or against. The reason for such a rule is that the support and against relationship is directional and can only be directed from factual statements to subjective feelings.

In the process of definition, this paper found that the more exciting speeches of the debaters sometimes contain obvious rhetorical devices to support their views, and the exciting speeches are often accompanied by the change of votes on both sides during the debate, so the label of “rhetorical devices” is added.

Moreover, each premise can only correspond to one argument (claim) during the marking process, because there is a clear pro-versus-versus relationship in the debate.

3.2. Design Principles

In this paper, we construct a Chinese argument corpus for argument mining and sentiment computing, using manual annotation. To achieve the above goals, corresponding strategies are needed to support: 1) determining the corpus acquisition strategy (oddly enough); 2) developing the corresponding annotation scheme and annotation specification; 3) designing the evaluation of annotation quality so as to confirm the usability of the corpus; 4) scaling the subsequent corpus.

3.3. Corpus Acquisition

Debate tournaments are a very common form of debate. Common and famous

debate tournaments include World Chinese Debate Championship, International Chinese Debate Invitational Tournament, Asia Pacific College Chinese Debate Tournament, Chinese Debate World Cup, and International Debate Invitational Tournament. The debate topics are often related to our life, such as “should college students be market-oriented” and “is this a face-saving society”. The debaters of debate tournaments often have richer argumentative logic and more complete argumentative structure, and more argumentative information can be obtained from their speeches, so this paper uses the data of debate tournaments as a corpus.

“QI PA SHUO” is a debate show produced in 2014. It aims to find the “most articulate people” in the Chinese language world with unique views and eloquence. The show will be held through Baidu, Zhihu, and Weibo. In the background, the program selects the most popular issues in the fields of people’s livelihood, humanities, emotions, life, business, entrepreneurship, etc. and invites netizens to participate in the survey and vote. In this paper, we take the debaters’ speeches in “QI PA SHUO” as the text information, and analyze and label the labeling scheme under the premise of ensuring the integrity of the argumentative components.

Current size of collection: As of the update progress of the episodes at the time of writing this article, a total of 21 episodes of the first season were collected. Annotated with sentence-level input and spanning 15 days.

3.4. Discourse Annotation

Corpus annotation is the process of processing the original corpus and selecting the annotated appendages and representations that facilitate the storage of the corpus for machine reading. TEI (text encoding initiative) is an international information encoding standard for machine readable discourse, which is researched and developed with the participation of many countries and has the characteristics of easy text description and suitable for multi-domain annotation and analysis. Currently, many large corpus resources including The British National Corpus are based on the TEI annotation guidelines.

In this paper, through the combination of TEI annotation and custom annotation, the input granularity and annotation granularity are both at the sentence level, and the basic framework of the Chinese argument corpus annotation system is as follows.

Argument Model = (theme, claim, premise, [instance], [modality], instance, emotion, category, [figures], debater, [votes], final vote, results)

The meaning of each label is as follows: theme, claim, argument, premise, instance, modality, stance, emotion, category, figures, debater, votes, final vote, results, debater, votes, final vote, and results. Examples, modal words, rhetorical devices and votes can be empty, but the rest of the items cannot be empty. In the debating process, there will be cases when debaters do not use examples or rhetorical techniques, so they can be marked as empty; while in the debating

process of “QI PA SHUO”, the moderator often does not give the current real-time vote count of both sides, so there is no way to know the real-time change of vote count in each round, i.e. the vote count can be marked as empty.

Example of labeling:

1) 肖骁: 我觉得一个人如果连自我包装和自我管理的能力都没有, 那你真的是失去了核心的竞争力, 累死我了让我休息一会!

[这是不是一个看脸的社会, theme]//辩论主题(theme)

[一个人如果连自我包装和自我管理的能力都没有, 那你真的是失去了核心的竞争力, premise]

[真的, modality]

[against]

[positive]

[NA]

[肖骁, debater]

[(51, 49), votes]

2) 姜思达: 其实钱是重要的吗, 别的也是重要的, 你今天能力也是重要的、你钱也是重要的, 但是如果这样我们就能得出这是一个看脸的社会, 那我们就能得出任何一个结论, 这是看钱的社会、这是看爹的社会、这是看身材的社会, 所以这样论述是没有什么道理的。

[这是不是一个看脸的社会, theme]//辩论主题(theme)

[你今天能力也是重要的、你钱也是重要的, instance]

[against]

[positive]

[PD]

[这是看钱的社会、这是看爹的社会、这是看身材的社会, figures]

[姜思达, debater]

[(51, 49), votes]

3.5. Quality Control

This subsection assesses the quality of the textual components of a multimodal argument corpus. This thesis uses the Krippendorff's α_U algorithm to assess the reliability of the annotation of the argument components. Krippendorff's α_U is able to take into account the differences in the annotation boundaries. In addition, the thesis uses the percentage agreement rate algorithm and two other chance-corrected agreement rate algorithms, Fleiss Kappa and Krippendorff's α . According to the statistics conducted in this thesis in the argument corpus, each debater's complete statement tends to each debater's complete statement often contains only one core argument (claim). These three agreeableness algorithms not only provide good approximations for assessing the reliability of annotation quality at the sentence level, but also enable quality comparisons with other corpora that use sentences as the annotation unit.

Table 2 represents the calculated results of the agreement test for the argumentative corpus. It can be seen that most of the agreement values are above 0.5,

Table 2. Result of consistency test.

Label	percentage agreement	Fleiss Kappa	Krippendorff's α	Krippendorff's α_U
Claim (argument/focus)	0.851	0.612	0.712	0.683
Premise (arguments/premises)	0.792	0.618	0.722	0.702
Instance	0.910	0.422	0.602	0.687
Modality (modal words)	0.774	0.534	0.467	0.355
Category (emotional category)	0.710	0.654	0.485	0.453
Figures (rhetorical device)	0.621	0.512	0.402	0.395

and all of them are above 0.7 except for rhetorical devices, which indicates that the inter-annotator agreement of the argumentative components is at a high level, and therefore the annotation quality is reliable.

After discussion, the reason why the percentage agreement of the label Figures is lower than 0.7 is that some rhetorical devices are used more obscurely during the debaters' speeches, "rhetorical questions" and "irony" are difficult to distinguish in many cases, and "puns" are difficult to judge accurately in the more complex debaters' speeches. It is difficult to distinguish between "rhetorical question" and "irony" in many cases, and it is difficult to accurately judge "pun" rhetorical techniques in the more complex debaters' speeches. Overall, it seems that although different rhetorical devices are clearly defined, the labeling process is still influenced by the subjective understanding of the debater's speech, resulting in a lower consistency test for this label than for other labels.

Table 3 shows the results of the consistency test for the label "sentiment polarity" of the argumentative corpus. From **Table 3**, it can be seen that the consistency of label "sentiment polarity" is generally high, indicating that the quality of labeling is reliable.

Sentiment polarity is an important label in sentiment computing, and sentiment polarity analysis is the process of analyzing, processing, generalizing, and mathematically modeling subjective texts with sentiment overtones. According to the different granularity of the input, sentiment classification can be classified into phrase-level, sentence-level, and chapter-level levels.

In affective polarity judgment, if the labels are classified as positive, negative, or neutral, then it can be classified as a triclassification problem. If there is no neutral text, it can be categorized as a dichotomous problem.

After the analysis of the sentiment polarity consistency test, the negative sentiment polarity consistency results were lower than the positive sentiment polarity as well as the neutral sentiment. With the knowledge of the corpus, it was

Table 3. Result of consistency test.

Label	percentage agreement	Fleiss Kappa	Krippendorff's α
Claim (positive)	0.942	0.821	0.712
Claim (negative)	0.712	0.693	0.682
Claim (null)	0.741	0.512	0.402

concluded that the statements made by the debaters based on their positions during the debate of “QI PA SHUO” tend to have positive affective polarity, and according to the rules of the debate, the number of votes brought by the debaters’ statements is an important factor in determining the outcome of the debate, so the debaters tend to argue their points on the basis of positive emotions. Negative emotional polarity is slightly more difficult to label than positive and unemotional polarity, and tends to be better labeled for debaters who are good at expressing emotions in arguments. When the emotional polarity is no-emotional polarity, the consistency test results are lower than the first two. The neutral emotional statements in the debate process mostly come from the debate moderator, who, after listening to both debaters’ statements, usually discusses their views from an objective perspective, mostly objective statements that do not contain emotions. The consistency test of the overall results is greater than 0.7.

Table 4 shows the results of the consistency test for the label “position” of the argumentative corpus. From **Table 4**, we can see that the consistency of the label “position” is generally high.

After the analysis of the consistency test of argumentative positions, the overall results of the consistency test reached a high level. The reason for this is related to the format of “QI PA SHUO”, which is based on the rules of traditional debating competitions, where the initial votes of both sides are counted and the running votes are calculated through each round of arguments. Since both sides of the debate have distinct positions, the “position” is in most cases the same as the side the debaters are on. However, compared to traditional debating competitions, “QI PA SHUO” has a higher degree of freedom and is accompanied by speeches from the moderator and guests, and in some cases, the speeches of the guests or moderator may be neutral or difficult for the marker to judge. The neutral statements are often accompanied by a slight shift in the subjective understanding of the markers. In general, the consistency test is greater than 0.9 for the debaters’ statements, which proves the quality of the dataset.

4. Corpus Applications

4.1. Expansion of the Argumentative Corpus

Existing Chinese argument corpora have corresponding limitations in their related research areas, such as user comments in user-generated texts, which are often only one or two sentences in length and thus cannot contain rich argumentative information, and the elements that can be extracted are very limited.

Table 4. Result of consistency test.

Label	percentage agreement	Fleiss Kappa	Krippendorff's α
Stance (Support)	0.985	0.921	0.902
Claim (Against)	0.912	0.893	0.882
Claim (Neutral)	0.741	0.612	0.702

After studying the text of “QI PA SHUO”, it is found that the excellent debaters in it have wonderful polemical speeches, which contain rich polemical elements that can extract not only the arguments with distinct positions, but also deeper information in them, such as rhetorical methods, which can help this paper determine whether there is a significant impact on the debate results when rhetorical techniques are added to the polemical text.

The corpus contained in this paper is the text of the first season of “QI PA SHUO”, and the corpus can continue to be expanded according to the updates of “QI PA SHUO” to obtain a larger and higher quality Chinese argumentative corpus.

4.2. Automatic Debate

Essay writing, philosophy books and articles can be useful in the direction of argument mining. English-oriented argumentative corpora often see the task of collecting students' argumentative essays and using them to create an argumentative corpus to explore argumentative structures. Early on Lawrence et al. also used argument mining techniques to annotate (Hua & Wang, 2018) the argumentative structures that appear in 19th century philosophy collections. In addition to collecting student essays, argument mining can also be applied in the education industry to build models to extract arguments from the collected essays and to score (Wachsmuth et al., 2016) student essays.

4.3. Reading Comprehension

Reading comprehension can also be seen as a branch of automatic debate, when the quality of the corpus is sufficient, then the core ideas in the text can be extracted precisely by combining it with algorithmic models. Especially today, with the development of the web, web users also provide a lot of available data for the direction of debate mining. For example, applications such as Zhihu, Weibo, etc. contain a lot of data containing argumentative text under specific topics. Ivan and Gurevych collected users' debates from online debate forums and annotated a data set containing 16,000 argument pairs through web crowdsourcing. Each argument pair consists of two arguments on the same topic, and the first argument in the pair is guaranteed to be richer in argumentative information than the second argument.

Therefore, more accurate judgment and extraction of argumentative arguments will undoubtedly help users to have a better understanding of the core ideas of the text.

5. Conclusion

Against the background of the gradual and urgent demand for corpus construction for argument mining, this paper constructs a Chinese argument corpus. The first stage of work and annotation has been completed. It contains a total of 21 debates of “QI PA SHUO”. This paper adds more novel tags such as rhetorical devices on the basis of the classical debate model, and helps this paper to better extract arguments and arguments at this stage and in the future work. In addition, because of the sharp pro and con sides in the debate and the sentiment polarity, this paper will try to combine with the sentiment classification task in the subsequent work.

The practical significance of the research done in this paper is to provide a more complete and diverse corpus for labeling in the direction of argument mining, and to provide reliable data for sentiment classification tasks as well, taking into account sentiment computation.

In addition, it is difficult to build the corpus flawlessly, and even if the labels are clearly defined, the context may still be difficult to understand in the annotation process, so we need to gradually expand the size of the corpus in the subsequent work to make this paper more mature. We will continue to improve and revise it in the next step.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- Burns, M. K. (2014). How to Establish Interrater Reliability. *Nursing*, *44*, 56-58. <https://doi.org/10.1097/01.NURSE.0000453705.41413.c6>
- Durmus, E., & Cardie, C. (2019). A Corpus for Modeling User and Language Effects in Argumentation on Online Debating. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the Annual 57th Meeting of the Association for Computational Linguistics* (pp. 602-607). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1057>
- Fleiss, J. L. (1971). Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin*, *76*, 378-382. <https://doi.org/10.1037/h0031619>
- Freeman, J. B. (1991). Appendix. Two Variations on the Standard Approach to Diagramming Arguments: Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *23*, 506-518.
- Hua, X., & Wang, L. (2018). Neural Argument Generation Augmented with Externally Retrieved Evidence. In I. Gurevych, & Y. Miyao (Eds.), *Proceedings of the Annual 56th Meeting of the Association for Computational Linguistics* (pp. 219-230). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1021>
- Jo, Y., Poddar, S., Jeon, B. et al. (2018). Attentive Interaction Model: Modeling Changes in View in Argumentation. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the Conference 2018 of the North American Chapter of the Association for Computational*

- Linguistics: Human Language Technologies* (No. 1, pp. 103-116). Association for Computational Linguistics. <https://doi.org/10.18653/v1/N18-1010>
- Krippendorff, K. (2004) Measuring the Reliability of Qualitative Text Analysis Data. *Quality & Quantity*, 38, 787-800. <https://doi.org/10.1007/s11135-004-8107-7>
- Krippendorff, K. (2007) *Computing Krippendorff's Alpha Reliability* (p. 43). Departmental Papers, Accounting Standards Council.
- Reed, C., Palau, R. M., Rowe, G. et al. (2008) Language Resources for Studying Argument. In *International Conference on Language Resources and Evaluation, LREC 2008* (pp. 2613-2618). European Language Resources Association.
- Toulmin, S. (1958). *The Uses of Argument*. Cambridge University Press.
- Wachsmuth, H., Al-Khatib, K., & Stein, B. (2016). Using Argument Mining to Assess the Argumentation Quality of Essays. In Y. Matsumoto, & R. Prasad (Eds.), *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers* (pp. 1680-1691). The COLING 2016 Organizing Committee.
- Walton, D. (2009). Argumentation Theory: A Very Short Introduction. In G. Simari, & I. Rahwan (Eds.), *Argumentation in Artificial Intelligence* (pp. 1-22). Springer. https://doi.org/10.1007/978-0-387-98197-0_1
- Xu, K. (2016). Research and Development of Argumentation Mining. *International Academic Trends*, No. 3, 7-8.