AntCritic: Argument Mining for Free-Form and Visually-Rich Financial Comments

Huadai Liu[1*], **Wenqiang Xu**[2*], **Xuan Lin**[2*], **Jingjing Huo**[2], **Hong Chen**[2], **Zhou Zhao**[1]
Zhejiang University[1], Ant Group[2]
liuhuadai@zju.edu.cn, {yugong.xwq, daxuan.lx,huojingjing.hjj}@antgroup.com
wuyi.ch@antgroup.com, zhaozhou@zju.edu.cn

Abstract

Argument mining aims to detect all possible argumentative components and identify their relationships automatically. As a thriving task in natural language processing, there has been a large amount of corpus for academic study and application development in this field. However, the research in this area is still constrained by the inherent limitations of existing datasets. Specifically, all the publicly available datasets are relatively small in scale, and few of them provide information from other modalities to facilitate the learning process. Moreover, the statements and expressions in these corpora are usually in a *compact* form, which restricts the generalization ability of models. To this end, we collect a novel dataset *AntCritic* to serve as a helpful complement to this area, which consists of about 10k free-form and visually-rich financial comments and supports both argument component detection and argument relation prediction tasks. Besides, to cope with the challenges brought by scenario expansion, we thoroughly explore the fine-grained relation prediction and structure reconstruction scheme and discuss the encoding mechanism for visual styles and layouts. On this basis, we design two simple but effective model architectures and conduct various experiments on this dataset to provide benchmark performances as a reference and verify the practicability of our proposed architecture. We release our data and code in this *link*, and this dataset follows CC BY-NC-ND 4.0 license.

Keywords: Argument Mining, Large-scale Corpus, Visually-rich Document Understanding

1. Introduction

With the rapid development of Internet technology, millions of opinions and thoughts are thus transmitted and stored on various web pages and applications, containing a great deal of research and application value. Among them, some complicated scenarios bring extra challenges, such as processing academic exchanges or movie reviews. To this end, the task of argument mining appears and arouses growing attention.

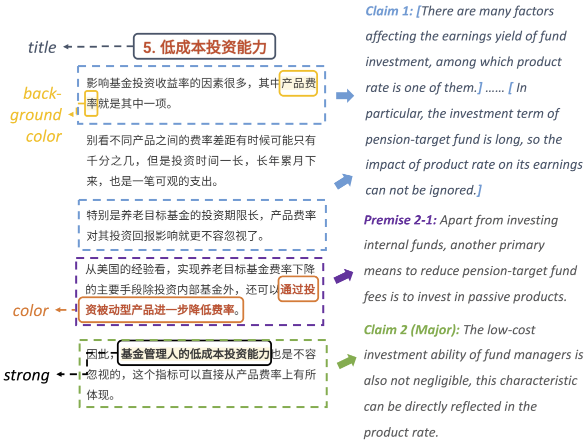
Given a persuasive article containing argumentative expressions, the task of argument mining aims to automatically identify all the potential argument components and correctly recover the corresponding argument structures. Unlike the other tasks, such as sentiment analysis or key-phrase extraction, this target requires the intelligent model to concurrently understand both semantic information and the logical structure and causal relationships within data. Despite the difficulties and challenges in the modeling process, the importance of this task is apparent. From the academic view, it can provide accurate information for the derivative subtasks (such as argument summarization and structured argument generation) and is conducive to understanding documents with complex semantic structures. While for the application aspect, the learned models can assist businesses or enterprises to automatically extract opinions and feedback or quickly trace the core opinions of prevailing topics without too much manual effort.

As a comprehensive and inclusive field, argument mining can be regarded as a combination of multiple interrelated sub-goals in essence. Therefore, to put more emphasis on academic analysis rather than constructing a complicated pipeline, researchers tend to discuss and address the issues only in a specific section, such as argument component detection (Gao et al., 2017; Daxenberger et al., 2017; Ruggeri et al., 2021), argument relation classification (Nguyen and Litman, 2016; Peldszus and Stede, 2015; Cocarascu and Toni, 2018), and stance detection (Wei et al., 2018; Ebrahimi et al., 2016; Johnson and Goldwasser, 2016). And to facilitate these subtasks, massive language resources (Reed et al., 2008; Habernal and Gurevych, 2017; Peldszus, 2015; Stab and Gurevych, 2017; Fergadis et al., 2021) are devoted to providing more effective and diversified data and supporting the learning process of models. Although the sufficient research and corpus mentioned above have established the foundation of this task, there is still a lack of further analysis for more variable and challenging scenarios. For example, the previous work tends to assume that all the arguments are expressed in a *compact* form, which means the statements of a single unit will be arranged consecutively. Thus non-adjacent clauses or sentences will be naturally regarded as different ones. But in practice, there might be some situations where semantic-consistent claims and

[*] Equal contributions

LREC-COLING 2024, pages 1306–1317
20-25 May, 2024. © 2024 ELRA Language Resource Association: CC BY-NC 4.0

premises get entwined with each other. Besides, although a considerable number of visually-rich documents are collected from social media such as Twitter (Bosc et al., 2016), Wikipedia (Biran and Rambow, 2011), and web blog posts (Habernal and Gurevych, 2017), the valuable auxiliary information from the other modalities is directly deserted in the construction of datasets. Furthermore, the existing corpora are relatively small in scale, and a large proportion of them do not support the subtask of relation prediction.



5. 低成本投资能力

影响基金投资收益率的因素很多，其中产品费率就是其中一项。

别看不同产品之间的费率差距有时候可能只有千分之几，但是投资时间一长，长年累月下来，也是一笔可观的支出。

特别是养老目标基金的投资期限长，产品费率对其投资回报影响就更不容忽视了。

从美国的经验看，实现养老目标基金费率下降的主要手段除投资内部基金外，还可以通过投资被动型产品进一步降低费率。

因此，基金管理人的低成本投资能力也是不容忽视的，这个指标可以直接从产品费率上有所体现。

Claim 1: [There are many factors affecting the earnings yield of fund investment, among which product rate is one of them.] ...... [In particular, the investment term of pension-target fund is long, so the impact of product rate on its earnings can not be ignored.]

Premise 2-1: Apart from investing internal funds, another primary means to reduce pension-target fund fees is to invest in passive products.

Claim 2 (Major): The low-cost investment ability of fund managers is also not negligible, this characteristic can be directly reflected in the product rate.

Figure 1: An example of visually-rich free-form documents in our proposed *AntCritic* dataset. Better viewed with color and zoom-in.

Considering these aspects mentioned above, we collect a novel dataset named *AntCritic* to facilitate the research of argument mining on visually-rich documents and free-form expressions. Concretely, we collect a total of about 10k argumentative financial comments from an open online forum supported by *Alipay*, which achieves the largest scale in this field as far as we know. In these documents, the visual patterns represented in the format of HTML tags and attributes are provided along with the text statements, and all the arguments are allowed to represent in a *free* form, i.e., a semantic-consistent argument might be scattered in non-adjacent segments, which is illustrated in Figure 1. This setting broadens the original definition of this problem and paves the way for possible future research. On this basis, to provide a preliminary solution to free-form argument mining, we propose a fine-grained setting as an extension to this field and explore different predictive components to tackle the corresponding problem. As for the fusion of visual patterns, we develop a style gate module to measure the effect of style attributes and utilize a joint encoding mechanism to represent the complex positions in the layout of documents. Moreover, we also develop different model variants to conduct in-depth analysis and discussions on the granularity of modeling units.

In conclusion, our contributions in this paper can be explained in the following three aspects.

- We collect and contribute a novel dataset *AntCritic* to argument mining. This corpus contains about 10k visually-rich free-form financial comments, which makes it the largest corpus in this area to our best knowledge.
- We thoroughly discuss the problem setting and the corresponding solution to free-form argument mining and explore a possible scheme to take advantage of auxiliary visual pattern information, which further broadens the scope of this field.
- We design two simple but effective model architectures and conduct extensive experiments to provide a reliable benchmark for our proposed dataset and serve as a preliminary solution to the task of argument mining on visually-rich free-form documents.

2. Related Work

2.1. Visually-rich Document Understanding

Depending on the accessibility of data organization, visually-rich documents can be roughly categorized into two groups. The first ones are *pattern-accessible* documents, such as webpages, and markdown files, in which all the visual patterns are controlled by clear scripts and can be explicitly accessed by analysts. On the contrary, in *pattern-inaccessible* documents like tickets and posters, there barely exist direct clues or instructions for the construction of documents, resulting in the vagueness of visual analysis for them. In this paper, we mainly discuss the former cases because the visual attributes represented by HTML tags are actually pattern-accessible ones.

Guided by explicit visual instructions, the rendering markups can greatly help analyze documents in a structured and human-like way. Wang et al. (2022) consider HTML tags and linguistic tokens as two individual sequences to model their relations in a cross-fusion way. Li et al. (2021); Deng et al. (2022) bring the self-supervised pre-training diagram into this field to drive models comprehensively to understand the whole webpage. Apart from these, some specialized architectures are also devised for some downstream tasks: Ashby and Weir (2020) regard HTML tags as auxiliary information to recognize named entities. Also, Chen et al. (2021) collect a novel question-answering dataset and formulate the task of structural reading comprehension on the web. Furthermore, Wang et al. (2021) design a novel attention-based framework to generate relational table representations in finer granularity. These works inspire us to boost the performance of argument mining with the help of visual and structural information.

2.2. Argument Mining

Argument mining aims to automatically detect and model all possible argument structures in the given documents or text fragments. In previous studies, researchers primarily focus on classifying argument components. Lawrence et al. (2014); Nguyen and Litman (2016) apply Latent Dirichlet Allocation (LDA) to learn in an unsupervised way. Madnani et al. (2012) employ Conditional Random Field (CRF) to make predictions based on some pre-defined lexical properties. Alhindi and Ghosh (2021); Zhang et al. (2022) leverage the prevailing BERT and Bi-LSTM modules to tackle this problem with the help of pre-training. Operating on the constituency tree, Ruggeri et al. (2021) extract semantic information in a graph-based way. Moreover, Daxenberger et al. (2017); Schulz et al. (2018) explore the transferability between different domains, and Gao et al. (2017); Kees et al. (2021) thoroughly discuss the feasibility of reinforcement learning and active learning in this field, respectively. Besides, Nguyen and Litman (2016); Peldszus and Stede (2015); Cocarascu and Toni (2018) also devise various relation modeling schemes to identify the relations between arguments, which further advance the identification of argumentative structure at a higher level.

Considering it is pretty essential to provide enough samples and annotations for machine inference and reasoning, some valuable datasets are also collected from various domains, covering the types of scientific papers(Mayer et al., 2020), persuasive essays(Stab and Gurevych, 2017), sustainable development policies(Fergadis et al., 2021), legal texts(Zhang et al., 2022) and so forth. Apart from these, Bosc et al. (2016); Liu et al. (2017); Habernal et al. (2018) also collect datasets containing argument structures for the recognition of emotion or opinion preferences in comments and reviews. Lately, Mestre et al. (2021) contribute a large-scale audio-text dataset, promoting the application of multi-modal analysis in this area. However, these datasets are limited in scales and expression forms, and there is still no adequate research to combine other information into the analysis, making our research a meaningful complement to this field.

3. Dataset

Based on the aforementioned considerations, we aim to contribute a large-scale corpus to this field, named *AntCritic*. This dataset contains about 10k Chinese financial comments collected from an open online forum supported by *Alipay*, which covers the themes of fund introductions, stock market analysis, investment advice, etc. Before collecting all the original data from the online open forum, we confirmed that the users had permanently and irrevocably licensed the rights to the published content to the platform by means of the user agreement. And we have obtained the right from the platform to use the data for academic research and public release.

3.1. Data Annotation

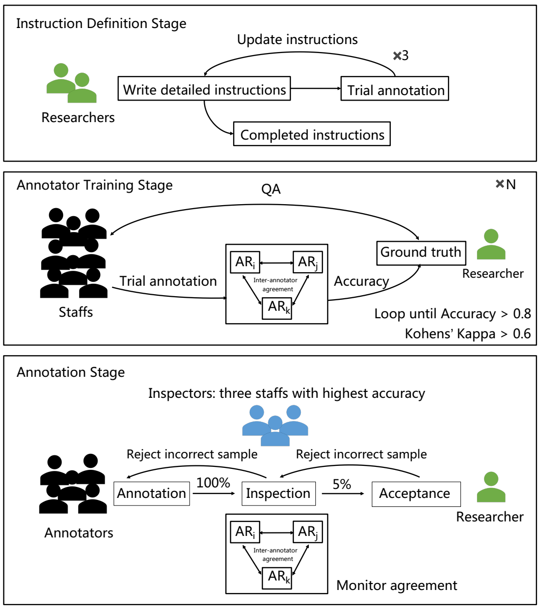


Figure 2: The overall annotation pipeline.

To ensure the quality of annotated data, we develop our annotation pipeline shown in Figure 2.

3.1.1. Instruction Definition Stage

In this stage, we will first draw up a series of detailed instructions, including the definition and format of different related terms. And then, we will carry out several rounds of trial annotations and refine the instructions so that the vagueness in instructions can be reduced to the minimum and the instructions are practical to all kinds of data. And here, we list some important terms as follows.

Table 1: Statistic Information Comparison of Datasets for Argument Mining

Dataset	Domain	#Docs	#Sents	#Claims	#Tokens	Unit	Rel?	Modal	Lang
Habernal and Gurevych	Web Discourse	340	3,899	211	84,817	Token	No	Text	EN
Mayer et al.	Scientific Papers	659	35,012	1,390	-	Token	Yes	Text	EN
Reed et al.	Various Genres	507	2,842	563	60,383	Clause	No	Text	EN
Stab and Gurevych	Persuaive Essays	402	7,116	2,108	147,271	Clause	Yes	Text	EN
Peldszus	Micro Texts	112	449	112	8,865	Segment	No	Text	EN
Fergadis et al.	SDG Policy	1,000	12,374	1,202	-	Sentence	No	Text	EN
Ours	Financial Comments	9,994	214,585	88,311	11,436,977	Segment	Yes	Text & HTML	CN

- A *claim* is defined as a group of argumentative segments that express a clear point of view, and the corresponding *premises* are defined as the fragments containing factual statements which underpin this opinion.

- There should be at least one claim in a document. The maximum of claims is set as 9, and the maximum of premises supporting the same claim is 4. It is also allowed to annotate some opinions as simple claims without any supporting premises.

- All the claims and premises should be semantically complete and individual. So the annotators may need to concatenate some non-adjacent segments to form a single integral argument. Besides, the annotators should also select a major claim that best matches the overall opinion.

3.1.2. Annotator Training Stage

After the refinement of instructions, eight staff with specific financial knowledge are employed to annotate the collected data. Before the formal annotation, we use some examples to help them get familiar with this task's goal and the corresponding requirements. Then, we will conduct rounds of trial annotations, evaluate the performance and statistics among all the annotators, and resolves the doubts of annotators in the QA phase. This process will repeat until the average accuracy reaches 80% and the Cohen's kappa (McHugh, 2012) agreement index is greater than 60%. Afterward, three annotators with the highest accuracy are selected as inspectors in the next formal annotation stage.

3.1.3. Annotation Stage

In the final stage, the formal annotation stage will be carried out in three steps: *Annotation-Inspection-Acceptance*. As mentioned above, a total of five annotators are required to label the data. And then, the other three inspectors check all the labeled results and reject the incorrect results. Finally, we sample 5% data to monitor the overall quality of annotation, and we accept the final annotations if and only if the agreement index and accuracy can reach the requirement.

3.2. Data Properties

To the best of our knowledge, our contributed *Ant-Critic* is the largest argument mining dataset in scale, and the detailed comparison can be found in Table 1. Besides, there are also some unique properties compared with other datasets proposed previously, which are depicted in Figure 1 and 3, and listed as follows. *i) Multi-Modality*: all the visual patterns, structural information and linguistic expressions can be accessed directly and explicitly. *ii) Discontinuity*: the text segments belonging to the same argument may be placed in a discontinuous manner. *iii) Non-Monotonicity*: some segments of different argumentative components may be arranged alternatively.

These characteristics described above primarily stem from the nature of the data source. Because there are no strict writing instructions for creators on the platform, the organization of documents may not be compact and precise enough, thus making the segments from different argument components interweave with each other. Although these properties significantly increase the learning difficulty, this setting reduces the requirements on the corpus to the minimum, which provides more opportunities for further research on free-form argument mining and improves the feasibility and generalization of learned models in practice.

3.3. Dataset Statistics

Table 2: Training/Validation/Test corpora statistics.

	Train	Validation	Test
# Comments	7,986	1,007	1,001
# Segments	433,342	40,296	33,913
# Claims	35,036	4,741	4,065
# Premises	37,022	4,102	3,788
# Claim Segment	69,823	10,225	8,263
# Premise Segment	127,379	13,620	11,321
# Character / Segment	22.55	22.10	22.84
# Segment / Claim	1.99	2.16	2.03
# Segment / Premise	3.44	3.32	2.99
% Font	16.85%	10.73%	18.72%
% Strong	9.98%	7.30%	10.10%
% Color	7.07%	6.03%	9.37%
% Blockquote	0.97%	0.62%	0.58%
% Supertalk	0.64%	0.69%	0.96%
% Header	0.24%	0.59%	0.78%

Table 2 demonstrates the statistic information of our proposed *AntCritic*. The upper part demonstrates the quantitative characteristics of texts, and the lower part shows the percentage of segments containing specified HTML tags.

4. Problem Definition

From an overall perspective, our goal of argument mining on visually-rich free-form documents can be decomposed into two targets, namely *Argument Component Detection* and *Argument Relation Prediction*, formulated as follows.

4.1. Argument Component Detection

As the crucial step in the task of argument mining, it aims to detect and classify all the potential argument components. Based on the annotation of this dataset, we only consider three possible component types, including *Claim*, *Premise* and *Non-Argument*. Apart from this, the confidence of arguments to be major claims should also get estimated to serve as a reference for human decision-making.

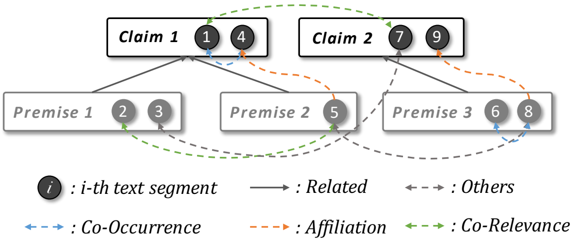


Figure 3: A simple diagram example for argument structures. Better viewed in color.

4.2. Argument Relation Prediction

To recover the argument structure in the documents, the relations between arguments should be understood correctly. However, the *discontinuity* and *non-monotonicity* of the dataset heavily impede us from slicing the entire document into multiple consecutive parts and predicting their relations. In light of this, we turn to recognize the relation types between individual segments and construct a more dense relation network within documents. Therefore we sort out all possible situations and accordingly define four relation types to cover them.

- **Affiliation:** The i-th segment is defined to *affiliate* with the j-th one if the latter belongs to a claim and the former is part of the premise supporting this claim. It maintains the vertical hierarchy of the entire document.
- **Co-Occurrence:** The i-th and j-th segments are defined to form a *Co-Occurrence* relation if they belong to the same argument. Without loss of generalization, every argumentative segment builds a *co-occurrence* relation with itself. It reveals the internal composition of a single argument component.
- **Co-Relevance:** The i-th and j-th segments are defined to form a *Co-Relevance* relation if they belong to the sibling components which support the same argument. And we assume that all the claims are trivially defined to support a virtual main topic. This kind of relationship can help recover the horizontal argumentative structure of the document.
- **Other:** The *Other* relation includes all the cases not covered by the definitions listed above.

Their diagram can also be found in Figure 3. It is straightforward to notice that there tend to be some redundancies in the relation annotations. These redundancies seem unnecessary, but they can enhance the robustness of relation predictions and act as an extra constraint to guide the model to consider the global argumentative structure and understand the relation types in depth.

5. Method

5.1. Model Architecture

This section will describe our proposed preliminary solution to the aforementioned problem setting on the *AntCritic* dataset. Given a structured webpage with various markups and attributes $\mathbf{D} = \{(\mathbf{d}_i, \mathbf{m}_i, p_i, s_i)\}_{i=1}^{n}$, our goal is to predict the labels of text segments $\mathbf{L} = \{l_i\}_{i=1}^{n}$, the confidence to be major claims $\{c_i\}_{i=1}^{n}$, and the relations between them $\mathbf{R} = \{\{r_{ij}\}_{i=1}^{n}\}_{j=1}^{n}$, where n is the total number of text segments and the quadruple $(\mathbf{d}_i, \mathbf{m}_i, p_i, s_i)$ represents the token sequence, style marks, paragraph position and segment position of the i-th text segment respectively. To quantify the impact of modeling granularities on the prediction, we construct the token-level and segment-level architectures to extract information and make predictions individually.

5.1.1. Token-Level Model

Because of the exceeding number of input tokens, we process each segment separately and only conduct the argument component detection task in this part. The overall pipeline of this model is demonstrated in the appendix. We consider both the character and word sequences of the input segments to investigate all the possible modeling schemas and further explore the dataset's properties. Concretely, we extract the aggregated representations via character-based and word-based backbones in a *First-Last-Avg* manner and finetune these models. Afterward, we apply a simple multi-layer perceptron to generate class probabilities. And following the approach proposed by Wei and Zou (2019), three kinds of strategies for text augmentation are also employed to enhance the robustness of predictions, namely *Random Mask*, *Random Swap* and *Random Repeat*. In this computing process, we leave out the HTML tags from the calculation because the marks for a single segment can only provide quite limited structural information and visual difference for the label prediction.

5.1.2. Segment-Level Model

In the initialization, we also adopt the *First-Last-Avg* mechanism to generate the segment-level features, but all the parameters in backbones are frozen for less computation resource requirements.

Afterward, we project these features into the same latent space, denoted as $\mathbf{f}_i^{\mathrm{c}}$ and $\mathbf{f}_i^{\mathrm{w}}$ for the character-base and word-based aggregations of the i-th segment, respectively. Next, considering that these two groups of embeddings are complementary semantic views of the input segments that emphasize different semantic aspects, we fuse these two representations into a single feature sequence. To be more specific, we apply a cross-gate module to calculate the fused representations adaptively. The formulae are given as

$$\mathbf{g}_i^{\mathrm{c}} = \sigma(\mathbf{W}^{\mathrm{c}}\mathbf{f}_i^{\mathrm{c}} + \mathbf{b}^{\mathrm{c}}), \quad \mathbf{g}_i^{\mathrm{w}} = \sigma(\mathbf{W}^{\mathrm{w}}\mathbf{f}_i^{\mathrm{w}} + \mathbf{b}^{\mathrm{w}}), \tag{1}$$

$$\mathbf{f}_i = \mathbf{W}^{\mathrm{o}} \cdot [\mathbf{g}_i^{\mathrm{c}} \cdot \mathbf{f}_i^{\mathrm{w}}; \mathbf{g}_i^{\mathrm{w}} \cdot \mathbf{f}_i^{\mathrm{c}}] + \mathbf{b}^{\mathrm{o}}, \tag{2}$$

where $\mathbf{g}_i^{*} \in (0,1)^d, \mathbf{f}_i^{*} \in \mathbb{R}^d$ and the variables $\mathbf{b}^{*} \in \mathbb{R}^d$, $\mathbf{W}^{\mathrm{w}}, \mathbf{W}^{\mathrm{c}} \in \mathbb{R}^{d\times d}$, $\mathbf{W}^{\mathrm{o}} \in \mathbb{R}^{d\times 2d}$ are all learnable parameters. $\sigma(\cdot)$ is the sigmoid function and $[;]$ is the concatenation operator.

Meanwhile, the visual pattern $\mathbf{m}_i$ and structural position information (p_i, s_i) are also encoded into a series of embeddings denoted as $\mathbf{v}_i$ and $\mathbf{e}_i$, which can be given by

$$\mathbf{e}_i = [\mathbf{E}^{\mathrm{p}}_{p_i}; \mathbf{E}^{\mathrm{s}}_{s_i}], \quad \mathbf{v}_i = \mathbf{E}^{\mathrm{m}}\mathbf{m}_i, \tag{3}$$

where $\mathbf{m}_i \in \{0,1\}^6$ is the stack of indicators for font, strong, color, blockquote, supertalk and header tags / attributes in the i-th segment, respectively. And $\mathbf{E}^{\mathrm{p}}, \mathbf{E}^{\mathrm{s}} \in \mathbb{R}^{N\times(d/2)}, \mathbf{E}^{\mathrm{m}} \in \mathbb{R}^{d\times 6}$ are the trainable weights of look-up tables.

Then, we utilize a style gate to dynamically control the density of semantic information according to the style appearances and reveal what the authors want to emphasize. The pattern-aware embedding $\{\hat{\mathbf{f}}_i\}_{i=1}^{n}$ will be consequently generated as

$$\mathbf{g}_i^{\mathrm{g}} = \sigma(\mathbf{W}^{\mathrm{g}}[\mathbf{v}_i; \mathbf{f}_i] + \mathbf{b}^{\mathrm{g}}), \quad \hat{\mathbf{f}}_i = (1 + \mathbf{g}_i^{\mathrm{g}})\mathbf{f}_i, \tag{4}$$

where $\mathbf{g}_i^{\mathrm{g}} \in (0,1)^d, \hat{\mathbf{f}}_i \in \mathbb{R}^d$, and $\mathbf{W}^{\mathrm{g}} \in \mathbb{R}^{d\times 2d}, \mathbf{b}^{\mathrm{g}} \in \mathbb{R}^d$ are learnable parameters.

And next, we need to utilize a sequence model to integrate the global information and form a series of context-aware representations given by

$$(\tilde{\mathbf{f}}_1, \ldots, \tilde{\mathbf{f}}_n) = \Omega((\hat{\mathbf{f}}_1 + \mathbf{e}_1), \ldots, (\hat{\mathbf{f}}_n + \mathbf{e}_n); \Theta), \tag{5}$$

where $\Omega(\cdot;\Theta)$ denotes the sequence modeling component. In practice, we select Transformer (Vaswani et al., 2017) and GRU (Chung et al., 2014) as typical choices in the experiment.

Consequently, context-aware representations will be used to detect all the arguments and recognize the relations between segment pairs. For argument component detection, we employ two linear layers to obtain the class probabilities $\mathbf{a}_i \in (0,1)^3$ and major confidence $c_i \in (0,1)$, given by

$$\mathbf{a}_i = \phi(\mathbf{W}^{\mathrm{a}}\tilde{\mathbf{f}}_i + \mathbf{b}^{\mathrm{a}}), \quad c_i = \sigma(\mathbf{W}^{\mathrm{m}}\tilde{\mathbf{f}}_i + b^{\mathrm{m}}), \tag{6}$$

where $\mathbf{W}^{\mathrm{a}} \in \mathbb{R}^{3\times d}, \mathbf{W}^{\mathrm{m}} \in \mathbb{R}^d, \mathbf{b}^{\mathrm{a}} \in \mathbb{R}^3, b^{\mathrm{m}} \in \mathbb{R}$ and $\phi(\cdot)$ is the softmax operator.

And for the task of argument relation prediction, we first project the features into the subspaces corresponding to the source and destination ends of relations, given as

$$\mathbf{f}_i^* = \mathbf{W}_2^*(\mathrm{ReLU}(\mathbf{W}_1^*\mathbf{f}_i + \mathbf{b}_1^*)) + \mathbf{b}_2^*,\ * \in \{s, d\}. \quad (7)$$

where the dimension of parameters keeps consistent with Equation 1. Next, two different pair-wise classification components are utilized to calculate the relation probabilities $\mathbf{r}_{ij} \in (0,1)^4$, which can be formulated as

- **Mul-Add:**

$$\mathbf{r}_{ij} = \phi(\mathbf{W}^{\mathrm{r}}[\mathbf{f}_i^{\mathbf{s}} + \mathbf{f}_j^{\mathbf{d}}; \mathbf{f}_i^{\mathbf{s}} \cdot \mathbf{f}_j^{\mathbf{d}}] + \mathbf{b}^{\mathrm{r}}), \quad (8)$$

- **Biaffine:**

$$\mathbf{r}_{ij} = \phi([\mathbf{f}_i^{\mathbf{s}}; 1]^{\mathsf{T}}\mathbf{U}[\mathbf{f}_j^{\mathbf{d}}; 1] + \mathbf{b}^{\mathbf{r}}), \quad (9)$$

where $\mathbf{U} \in \mathbb{R}^{(d+1)\times 4 \times (d+1)}, \mathbf{W}^{\mathrm{r}} \in \mathbb{R}^{4\times 2d}, \mathbf{b}^{\mathrm{r}} \in \mathbb{R}^d$ are trainable parameters. And the detailed calculation mechanism of the segment-level model is shown in the appendix.

5.2. Training and Inference

As illustrated in the descriptions in the sections above, the overall calculation can be conducted end-to-end. Considering the entire task is composed of three subtasks (including major confidence estimation), we optimize the model by combining these three objectives, formulated as follows.

Argument Component Detection The subtask of argument component detection is essentially a multi-classification problem, so we apply the cross-entropy loss function to tackle this.

$$\mathcal{L}_c = -\sum_{i=1}^{n}\sum_{j=0}^{2}\mathbb{I}(\hat{l}_i = j)\log((\mathbf{a}_i)_j), \quad (10)$$

where the ground-truth component label $\hat{l}_i$ equals to 0, 1, 2 if the i-th segment is annotated as *non-argument*, *claim* and *premise*, respectively.

Argument Relation Prediction Similarly, the optimization constraint of argument relation prediction can be given by

$$\mathcal{L}_r = -\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=0}^{3}\mathbb{I}(\hat{r}_{ij} = k)\log((\mathbf{r}_{ij})_k) \quad (11)$$

where the ground-truth relation label $\hat{r}_{ij}$ is assigned as 0, 1, 2, 3 where the relation between i-th and j-th segments is considered as *other*, *affiliation*, *co-occurrence* and *co-relevance*, respectively.

Major Confidence Estimation The estimation of major confidence will be approximated as a binary classification task and get solved by

$$\mathcal{L}_m = -\sum_{i=1}^{n}(\hat{m}_i\log(c_i) + (1-\hat{m}_i)\log(1-c_i)), \quad (12)$$

where $\hat{m}_i = 1$ if the i-th segment is annotated as a part of the major claim, otherwise $\hat{m}_i = 0$.

Finally, the segment-level model will be optimized in a multi-task manner, and the overall loss function is given by

$$\mathcal{L} = \lambda_c\mathcal{L}_c + \lambda_r\mathcal{L}_r + \lambda_m\mathcal{L}_m, \quad (13)$$

where λ_* are the balancing hyper-parameters. As for the token-level model, we only calculate the first term because the other two subtasks are not conducted on this.

While in the inference, we directly output the major confidence c_i and choose the labels with the highest probabilities as the prediction results.

6. Experiment

6.1. Metrics

We adopt three trustworthy criteria to evaluate our proposed solutions, namely *Component F1 Score*, *Relation F1 Score* and *Major Density*. For the Component & Relation F1 Scores, we calculate the *micro*-, *macro*- and *weighted*-F1 scores on the results of argument component detection and relation prediction. It's worth noting that the *Other* type of relation will not be considered in the metric calculation because this type does not mean that there is *no* relationship between the corresponding segments. As for the *Major Density*, we choose to normalize all the confidences of claim segments and sum up the scores corresponding to the ground-truth major claims, which can be given by

$$M = \frac{\sum_{i=1}^{n}(\tilde{c}_i \cdot \mathbb{I}(\hat{m}_i = 1))}{\sum_{j=1}^{n}\tilde{c}_j} \quad (14)$$

in which $\tilde{c}_i = c_i$ when $\hat{l}_i = 1$ and otherwise $\tilde{c}_i = 0$. In this mechanism, we can quantify the relative confidence density of major claims without being affected by the ratio of major segments.

6.2. Comparison and Analysis

In this section, we conduct a series of experiments to estimate the performance of the token-level and segment-level models and provide benchmark results on our proposed *AntCritic* dataset.

Table 3: Results of different settings for segment-level models. The best results are given in **bold**.

Row	Module	Relation Modeling	Using HTML?	Component Detection				Relation Prediction		
				Mac.	Mic.	Weig.	Major	Mac.	Mic.	Weig.
1	MLP	Biaffine	Yes	63.24	66.68	66.35	43.27	20.26	26.05	20.64
2	MLP	Mul-Add	Yes	62.59	66.07	65.75	42.29	22.61	27.18	22.93
3	TransFM	Biaffine	Yes	65.78	69.01	68.79	47.49	21.64	26.34	22.27
4	TransFM	Mul-Add	No	65.16	68.49	68.10	46.01	21.60	26.55	22.94
5	TransFM	Mul-Add	Yes	65.24	68.67	68.28	46.30	**23.84**	28.58	24.11
6	Bi-GRU	Biaffine	Yes	**67.23**	**70.22**	**70.02**	48.75	21.66	26.27	21.68
7	Bi-GRU	Mul-Add	No	66.24	69.72	69.16	49.28	22.00	27.55	23.03
8	Bi-GRU	Mul-Add	Yes	67.05	70.00	69.84	**49.88**	23.49	**29.84**	**24.77**

6.2.1. Analysis of Token-Level Model

The performances of different settings for the token-level model are listed in Table 4. The first two rows display the performances of word-based and character-based pipelines, respectively. And in the *Ensemble* setting, the final outputs are given as the average results of these two streams. From the table, we can find that the *Ensemble* setting outweighs the others in general, and the word-based stream is inferior to the character-based one. This can be explained in the following three folds. First, the character-based backbone is trained on the financial corpus, and the word-based one is trained on the data of Wikipedia and news commentaries. There will be a larger domain and semantic gap for the word-based backbone to fit our proposed dataset. Second, the word vocabulary is much bigger than the character one, making the token embeddings of some low-frequency words not well-tuned. Besides, the ensemble strategy can reduce the vagueness and uncertainty of probabilities and enhance the stability of predictions.

Table 4: Results of different settings for token-level models. The best results are given in **bold**.

Setting	Macro	Micro	Weighted
Word	54.30	61.64	58.73
Character	57.62	63.39	61.56
Ensemble	**58.32**	**63.87**	**62.15**

6.2.2. Analysis of Segment-Level Model

Table 3 demonstrates the evaluation results corresponding to different components and input modalities choices. And we try to analyze the performances in the following three aspects.

Impact of Modal Inputs As shown in Row 4,5,7,8 of Table 3, it is evident that the visual patterns and structural information indeed improve the performances of all these tasks to different extents, which implies that the authors and creators tend to utilize some special attributes and layouts to emphasize their opinions and the corresponding expressions, and our proposed gating mechanism can capture this auxiliary information and help the model to focus on both semantic information and visual appearances of segments.

Choice of Sequence Modeling Components Both Row 1,3,6, and Row 2,5,8 illustrate the effects of different sequential modules. It is explicit that the Bi-GRU module is superior to the Transformer, and both of them achieve better performances than the simple MLP. We intuitively speculate that the range of dependency modeling may lead to the performance gap between different modules. Considering the related claims and premises are usually close in position, we infer that the Bi-GRU module keeps semantic dependencies within a relatively short term, which provides enough information for message passing and aggregation and prevents long-term noise in the calculation.

Choice of Relation Predictors Comparing Row 1,3,6 with Row 2,5,8, we can find that the choices of relation modules bring a noticeable impact on relation prediction, and slightly affect the performance on argument component detection. The *Biaffine* method generally behaves worse than the *Mul-Add* strategy. The reason can be inferred that the inner-product of the biaffine module mixes the information along the embedding dimension, which may smooth the differences within it. However, the operation of *Mul-Add* individually models the weights of different positions and maintains the discrepancy information on this dimension.

6.2.3. Comparison between Different Models

Comprehensively comparing the evaluation results shown in Table 3 and 4, we can notice that the interaction between segments brings a more significant effect than the dependency modeling within a single segment. This result aligns with the intuition that the context semantic information is more important than the inner structure of expression for the reasoning and detecting argument components.

7. Conclusion

In this paper, we discuss a meaningful problem extension in the field of argument mining, i.e., the processing of visually-rich free-form documents, and explore the corresponding preliminary solutions to utilize auxiliary visual information and generate free-form fine-grained predictions. Besides, we further collect and contribute a large-scale corpus *AntCritic* to facilitate this setting. The comprehensive experiments verify the reliability of this dataset and the feasibility of our proposed architectures.

Acknowledgements

This work was supported by the Ant Group Research Fund.

8. Bibliographical References

Tariq Alhindi and Debanjan Ghosh. 2021. “sharks are not the threat humans are”: Argument component segmentation in school student essays. In *BEA*.

Colin Ashby and David Weir. 2020. Leveraging html in free text web named entity recognition. In *COLING*.

Or Biran and Owen Rambow. 2011. Identifying justifications in written dialogs by classifying text as argumentative. *Int. J. Semantic Comput.*, 5:363–381.

Tom Bosc, Elena Cabrio, and Serena Villata. 2016. Dart: a dataset of arguments and their relations on twitter. In *LREC*.

Lu Chen, Xingyu Chen, Zihan Zhao, Danyang Zhang, Jiabao Ji, Ao Luo, Yuxuan Xiong, and Kai Yu. 2021. Websrc: A dataset for web-based structural reading comprehension. In *EMNLP*.

Junyoung Chung, Çaglar Gülçehre, Kyunghyun Cho, and Yoshua Bengio. 2014. Empirical evaluation of gated recurrent neural networks on sequence modeling. *ArXiv*, abs/1412.3555.

Oana Cocarascu and Francesca Toni. 2018. Combining deep learning and argumentative reasoning for the analysis of social media textual content using small data sets. *Computational Linguistics*, 44:833–858.

Johannes Daxenberger, Steffen Eger, Ivan Habernal, Christian Stab, and Iryna Gurevych. 2017. What is the essence of a claim? cross-domain claim identification. In *EMNLP*.

Xiang Deng, Prashant Shiralkar, Colin Lockard, Binxuan Huang, and Huan Sun. 2022. Dom-lm: Learning generalizable representations for html documents. *ArXiv*, abs/2201.10608.

J. Ebrahimi, Dejing Dou, and Daniel Lowd. 2016. Weakly supervised tweet stance classification by relational bootstrapping. In *EMNLP*.

Aris Fergadis, Dimitris Pappas, Antonia Karamolegkou, and Haris Papageorgiou. 2021. Argumentation mining in scientific literature for sustainable development. In *ARGMINING*.

Yang Gao, Hao Wang, Chen Zhang, and Wei Wang. 2017. Reinforcement learning based argument component detection. *ArXiv*, abs/1702.06239.

Ivan Habernal and Iryna Gurevych. 2017. Argumentation mining in user-generated web discourse. *Computational Linguistics*, 43:125–179.

Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. 2018. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *NAACL*.

Kristen Marie Johnson and Dan Goldwasser. 2016. Identifying stance by analyzing political discourse on twitter. In *NLP+CSS@EMNLP*.

Nataliia Kees, Michael Fromm, Evgeniy Faerman, and T. Seidl. 2021. Active learning for argument strength estimation. *ArXiv*, abs/2109.11319.

John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft. 2014. Mining arguments from 19th century philosophical texts using topic based modelling. In *ArgMining@ACL*.

Junlong Li, Yiheng Xu, Lei Cui, and Furu Wei. 2021. Markuplm: Pre-training of text and markup language for visually-rich document understanding. *ArXiv*, abs/2110.08518.

Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, and Hao Wang. 2017. Using argument-based features to predict and analyse review helpfulness. In *EMNLP*.

Ilya Loshchilov and Frank Hutter. 2017. Sgdr: Stochastic gradient descent with warm restarts. *arXiv: Learning*.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *ICLR*.

Value Simplex Technology Co. Ltd. 2020. Finbert. https://github.com/valuesimplex/FinBERT.

Nitin Madnani, Michael Heilman, Joel R. Tetreault, and Martin Chodorow. 2012. Identifying high-level organizational elements in argumentative discourse. In *NAACL*.

Tobias Mayer, Elena Cabrio, and Serena Villata. 2020. Transformer-based argument mining for healthcare applications. In *ECAI*.

Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.

Rafael Mestre, Razvan Milicin, S. Middleton, Matt Ryan, Jiatong Zhu, and T. Norman. 2021. M-arg: Multimodal argument mining dataset for political debates with audio and transcripts. In *EMNLP 2021*.

Huy V. Nguyen and Diane J. Litman. 2016. Context-aware argumentative relation mining. In *ACL*.

Andreas Peldszus. 2015. An annotated corpus of argumentative microtexts.

Andreas Peldszus and Manfred Stede. 2015. Towards detecting counter-considerations in text. In *ArgMining@HLT-NAACL*.

Sashank J. Reddi, Satyen Kale, and Sanjiv Kumar. 2018. On the convergence of adam and beyond. *ArXiv*, abs/1904.09237.

Chris Reed, Raquel Mochales Palau, Glenn Rowe, and Marie-Francine Moens. 2008. Language resources for studying argument. In *LREC*.

Nils Reimers and Iryna Gurevych. 2020. Making monolingual sentence embeddings multilingual using knowledge distillation. In *EMNLP*.

Federico Ruggeri, Marco Lippi, and Paolo Torroni. 2021. Tree-constrained graph neural networks for argument mining. *ArXiv*, abs/2110.00124.

Claudia Schulz, Steffen Eger, Johannes Daxenberger, Tobias Kahse, and Iryna Gurevych. 2018. Multi-task learning for argumentation mining in low-resource settings. In *NAACL*.

Christian Stab and Iryna Gurevych. 2017. Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43:619–659.

Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *ArXiv*, abs/1706.03762.

Daheng Wang, Prashant Shiralkar, Colin Lockard, Binxuan Huang, Xin Dong, and Meng Jiang. 2021. Tcn: Table convolutional network for web table interpretation. *Proceedings of the Web Conference 2021*.

Qifan Wang, Yi Fang, Anirudh Ravula, Fuli Feng, Xiaojun Quan, and Dongfang Liu. 2022. Webformer: The web-page transformer for structure information extraction. *ArXiv*, abs/2202.00217.

Jason Wei and Kai Zou. 2019. Eda: Easy data augmentation techniques for boosting performance on text classification tasks. *ArXiv*, abs/1901.11196.

Penghui Wei, Junjie Lin, and Wenji Mao. 2018. Multi-target stance detection via a dynamic memory-augmented network. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*.

Gechuan Zhang, Paul Nulty, and D. Lillis. 2022. Enhancing legal argument mining with domain pre-training and neural networks. *ArXiv*, abs/2202.13457.

Pierre Zweigenbaum, Serge Sharoff, and Reinhard Rapp. 2017. Overview of the second bucc shared task: Spotting parallel sentences in comparable corpora. In *BUCC@ACL*.

A. Implementation Details

We employ the *Sentence-BERT* architecture proposed by Reimers and Gurevych (2020) as the word-based backbone. It is constructed on the basis of BERT-like architectures and gets pre-trained on the dataset of BUCC mining task (Zweigenbaum et al., 2017), which is composed of Wikipedia articles and news commentaries. And the character-based backbone is adapted from *FinBERT* (Ltd, 2020), an open-source Chinese BERT model released by the AI lab of Value Simplex. The corpus used to pre-train this model is collected from the finance domain, including financial news, research reports & announcements, and financial Wikipedia entries. The maximal number of segments N in a single document is set to be 400. Except for the parameters in the pre-defined structures, the dimension d of intermediate representations and learnable parameters are set to be 384. The number of layers in the Transformer and GRU components is 3, the number of heads in the attention mechanism is 4, and the layer number of the final MLP-based label predictor is set as 3. In the training process, the maximum of learning rate is set to be 1e-4, and the balancing hyper-parameters $\lambda_c, \lambda_r, \lambda_m$ are set as 1, 1 and 0.5, respectively. For both token-level and segment-level models,

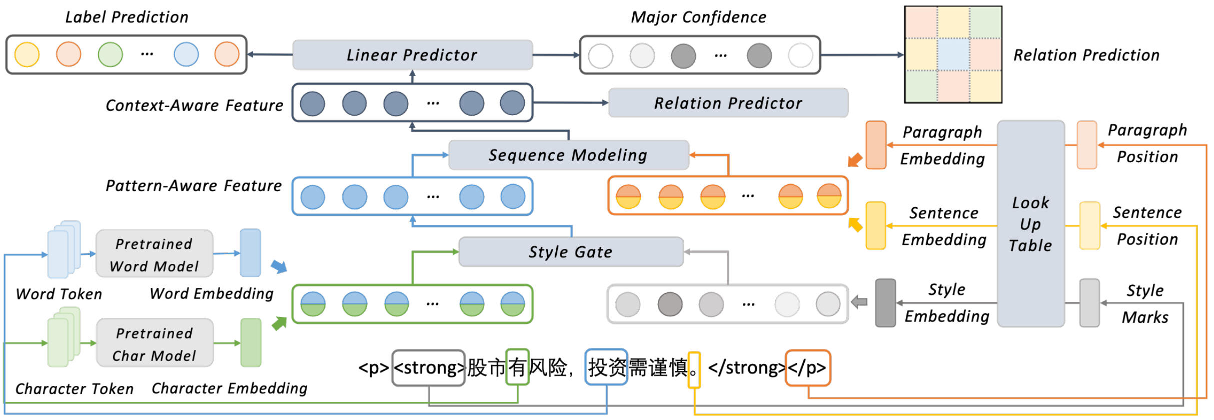


Figure 4: Overall diagram of segment-level model. Better viewed in color.

the learning rate is tuned with warm-up mechanism (Vaswani et al., 2017) and cosine annealing strategy (Loshchilov and Hutter, 2017) where the warm-up epoch is set as 1. To prevent overfitting, a dropout strategy with $p = 0.4$ is applied in the model. The training will last for 15 epochs and we select the checkpoint with the best performance on the validation dataset. As for the optimizer configuration, the AdamW (Loshchilov and Hutter, 2019; Reddi et al., 2018) optimizer with weight decay of 5e-5 is employed. The batch size is fixed to be 16. Apart from these, the probability of data augmentation for the token-level model is set as 0.3 and the ratio of modified tokens is 0.15. And in the overall experiment, the random seed is fixed and the details can be referred in the code files.

B. Limitation and Potential Risks

This section will discuss the limitations and potential risks related to our proposed dataset and architecture. In the overall view, we extend the original problem of argument mining and contribute a novel dataset to facilitate this setting. But because there is a lack of such discussion in the prior works, we fail to conduct some transfer/transplant experiments in this problem setting. Besides, this dataset is constructed in Mandarin, which may impede the researchers or developers in other countries from understanding and utilizing this corpus. As for the potential risks, the inherent bias in data collection may affect the further development of this dataset.

C. Architecture

We attach our segment-level model and token-level model architecture in Figure 4 and 5.

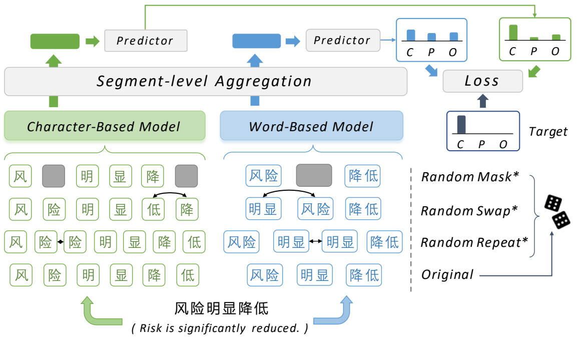
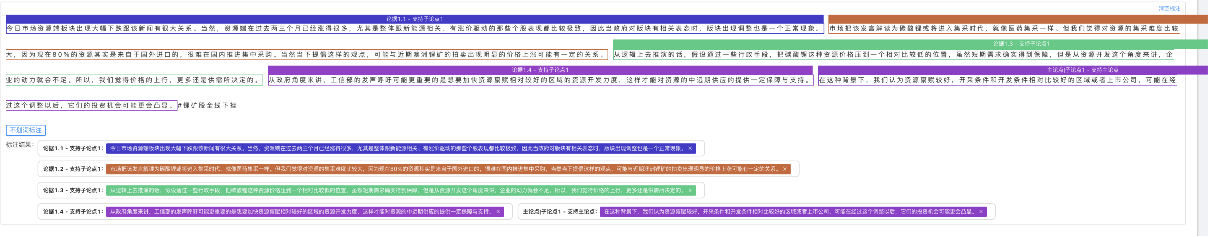


Figure 5: Overall diagram of the token-level model.

D. More Dataset Samples

We attach more visual-rich free-form documents examples here.



今日市场资源端板块出现大幅下跌跟该新闻有很大关系。当然，资源端在过去两三个月已经涨得很多，尤其是整体跟新能源相关、有涨价驱动的那些个股表现都比较极致，因此当政府对版块有相关表态时，版块出现调整也是一个正常现象。市场把该发言解读为碳酸锂或将进入集采时代，就像医药集采一样。但我们觉得对资源的集采难度比较大，因为现在80%的资源其实是来自于国外进口的，很难在国内推进集中采购。当然当下提倡这样的观点，可能与近期澳洲锂矿的拍卖出现明显的价格上涨可能有一定的关系。从逻辑上去推演的话，假设通过一些行政手段，把碳酸锂这种资源价格压到一个相对比较低的位置，虽然短期需求确实得到保障，但是从资源开发这个角度来讲，企业的动力就会不足。所以，我们觉得价格的上行，更多还是供需所决定的。从政府角度来讲，工信部的发声呼吁可能更重要的是想要加快资源禀赋相对较好的区域的资源开发力度，这样才能对资源的中远期供应的提供一定保障与支持。在这种背景下，我们认为资源禀赋较好，开采条件和开发条件相对比较好的区域或者上市公司，可能在经过这个调整以后，它们的投资机会可能更会凸显。#锂矿股全线下挫

标注结果：

论据1.1 - 支持子论点1：今日市场资源端板块出现大幅下跌跟该新闻有很大关系。当然，资源端在过去两三个月已经涨得很多，尤其是整体跟新能源相关、有涨价驱动的那些个股表现都比较极致，因此当政府对版块有相关表态时，版块出现调整也是一个正常现象。×

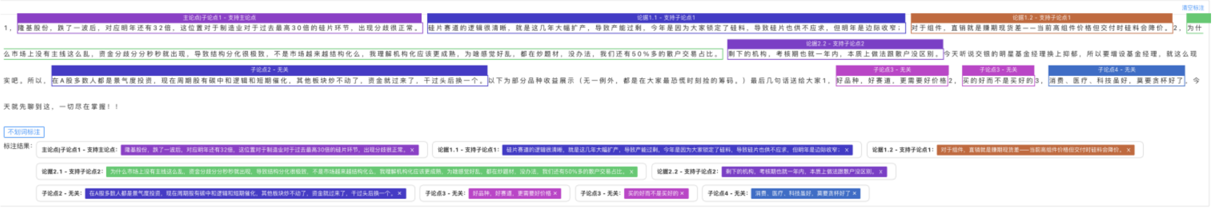
论据1.2 - 支持子论点1：市场把该发言解读为碳酸锂或将进入集采时代，就像医药集采一样。但我们觉得对资源的集采难度比较大，因为现在80%的资源其实是来自于国外进口的，很难在国内推进集中采购。当然当下提倡这样的观点，可能与近期澳洲锂矿的拍卖出现明显的价格上涨可能有一定的关系。×

论据1.3 - 支持子论点1：从逻辑上去推演的话，假设通过一些行政手段，把碳酸锂这种资源价格压到一个相对比较低的位置，虽然短期需求确实得到保障，但是从资源开发这个角度来讲，企业的动力就会不足。所以，我们觉得价格的上行，更多还是供需所决定的。×

论据1.4 - 支持子论点1：从政府角度来讲，工信部的发声呼吁可能更重要的是想要加快资源禀赋相对较好的区域的资源开发力度，这样才能对资源的中远期供应的提供一定保障与支持。×

主论点|子论点1 - 支持主论点：在这种背景下，我们认为资源禀赋较好，开采条件和开发条件相对比较好的区域或者上市公司，可能在经过这个调整以后，它们的投资机会可能更会凸显。×

Figure 6: Example 1



1，隆基股份，跌了一波后，对应明年还有32倍，这位置对于制造业对于过去最高30倍的硅片环节，出现分歧很正常。硅片赛道的逻辑很清晰，就是这几年大幅扩产，导致产能过剩，今年是因为大家锁定了硅料，导致硅片也供不应求，但明年是边际收窄；对于组件，直销就是赚期现货差——当前高组件价格但交付时硅料会降价。2，为什么市场上没有主线这么乱，资金分歧分分秒秒就出现，导致结构分化很极致，不是市场越来越结构化么，我理解机构化应该更成熟，为啥感觉好乱，都在炒题材，没办法，我们还有50%多的散户交易占比。剩下的机构，考核期也就一年内，本质上做法跟散户没区别。今天听说交银的明星基金经理换上抑郁，所以要增设基金经理，就这么现实吧。所以，在A股多数人都是景气度投资，现在周期股有碳中和逻辑和短期催化，其他板块炒不动了，资金就过来了，干过头后换一个。以下为部分品种收益展示（无一例外，都是在大家最恐慌时刻捡的筹码。）最后几句话送给大家1，好品种，好赛道，更需要好价格2，买的好而不是买好的3，消费、医疗、科技虽好，莫要贪杯好了，今天就先聊到这，一切尽在掌握！！

标注结果：

主论点|子论点1 - 支持主论点：隆基股份，跌了一波后，对应明年还有32倍，这位置对于制造业对于过去最高30倍的硅片环节，出现分歧很正常。×

论据1.1 - 支持子论点1：硅片赛道的逻辑很清晰，就是这几年大幅扩产，导致产能过剩，今年是因为大家锁定了硅料，导致硅片也供不应求，但明年是边际收窄；×

论据1.2 - 支持子论点1：对于组件，直销就是赚期现货差——当前高组件价格但交付时硅料会降价。×

论据2.1 - 支持子论点2：为什么市场上没有主线这么乱，资金分歧分分秒秒就出现，导致结构分化很极致，不是市场越来越结构化么，我理解机构化应该更成熟，为啥感觉好乱，都在炒题材，没办法，我们还有50%多的散户交易占比。×

论据2.2 - 支持子论点2：剩下的机构，考核期也就一年内，本质上做法跟散户没区别。×

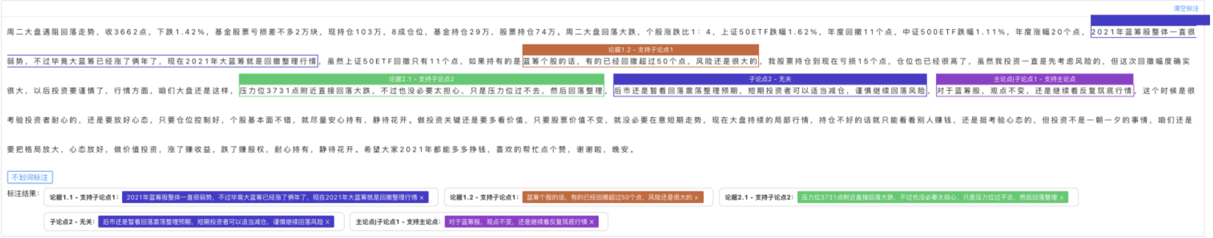
子论点2 - 无关：在A股多数人都是景气度投资，现在周期股有碳中和逻辑和短期催化，其他板块炒不动了，资金就过来了，干过头后换一个。×

子论点3 - 无关：好品种，好赛道，更需要好价格 ×

子论点3 - 无关：买的好而不是买好的 ×

子论点4 - 无关：消费、医疗、科技虽好，莫要贪杯好了 ×

Figure 7: Example 2



周二大盘遇阻回落走势，收3662点，下跌1.42%，基金股票亏损差不多2万块，现持仓103万，8成仓位，基金持仓29万，股票持仓74万。周二大盘回落大跌，个股涨跌比1：4，上证50ETF跌幅1.62%，年度回撤11个点，中证500ETF跌幅1.11%，年度涨幅20个点，2021年蓝筹股整体一直很弱势，不过毕竟大蓝筹已经涨了俩年了，现在2021年大蓝筹就是回撤整理行情，虽然上证50ETF回撤只有11个点，如果持有的是蓝筹个股的话，有的已经回撤超过50个点，风险还是很大的，我股票持仓到现在亏损15个点，仓位也已经很高了，虽然我投资一直是先考虑风险的，但这次回撤幅度确实很大，以后投资要谨慎了，行情方面，咱们大盘还是这样，压力位3731点附近直接回落大跌，不过也没必要太担心，只是压力位过不去，然后回落整理，后市还是暂看回落震荡整理预期，短期投资者可以适当减仓，谨慎继续回落风险，对于蓝筹股，观点不变，还是继续看反复筑底行情，这个时候是很考验投资者耐心的，还是要放好心态，只要仓位控制好，个股基本面不错，就尽量安心持有，静待花开。做投资关键还是要多看价值，只要股票价值不变，就没必要在意短期走势，现在大盘持续的局部行情，持仓不好的话就只能看着别人赚钱，还是挺考验心态的，但投资不是一朝一夕的事情，咱们还是要把格局放大，心态放好，做价值投资，涨了赚收益，跌了赚股权，耐心持有，静待花开。希望大家2021年都能多多挣钱，喜欢的帮忙点个赞，谢谢啦，晚安。

标注结果：

论据1.1 - 支持子论点1：2021年蓝筹股整体一直很弱势，不过毕竟大蓝筹已经涨了俩年了，现在2021年大蓝筹就是回撤整理行情 ×

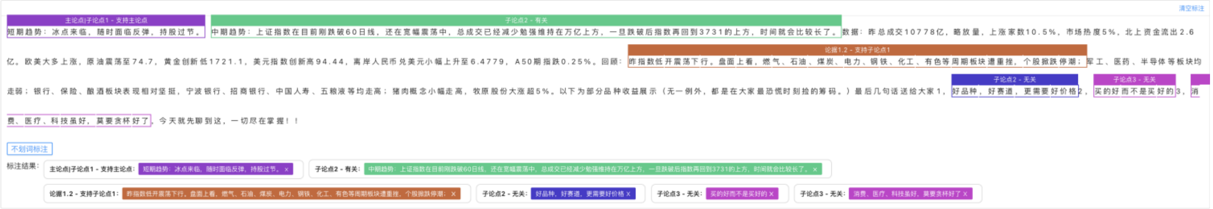
论据1.2 - 支持子论点1：蓝筹个股的话，有的已经回撤超过50个点，风险还是很大的 ×

论据2.1 - 支持子论点2：压力位3731点附近直接回落大跌，不过也没必要太担心，只是压力位过不去，然后回落整理 ×

子论点2 - 无关：后市还是暂看回落震荡整理预期，短期投资者可以适当减仓，谨慎继续回落风险 ×

主论点|子论点1 - 支持主论点：对于蓝筹股，观点不变，还是继续看反复筑底行情 ×

Figure 8: Example 3



短期趋势：冰点来临，随时面临反弹，持股过节。中期趋势：上证指数在目前刚跌破60日线，还在宽幅震荡中，总成交已经减少勉强维持在万亿上方，一旦跌破后指数再回到3731的上方，时间就会比较长了。数据：昨总成交10778亿，略放量，上涨家数10.5%，市场热度5%，北上资金流出2.6亿。欧美大多上涨，原油震荡至74.7，黄金创新低1721.1，美元指数创新高94.44，离岸人民币兑美元小幅上升至6.4779，A50期指跌0.25%。回顾：昨指数低开震荡下行。盘面上看，燃气、石油、煤炭、电力、钢铁、化工、有色等周期板块遭重挫，个股掀跌停潮；军工、医药、半导体等板块均走弱；银行、保险、酿酒板块表现相对坚挺，宁波银行、招商银行、中国人寿、五粮液等均走高；猪肉概念小幅走高，牧原股份大涨超5%。以下为部分品种收益展示（无一例外，都是在大家最恐慌时刻捡的筹码。）最后几句话送给大家1，好品种，好赛道，更需要好价格2，买的好而不是买好的3，消费、医疗、科技虽好，莫要贪杯好了，今天就先聊到这，一切尽在掌握！！

标注结果：

主论点|子论点1 - 支持主论点：短期趋势：冰点来临，随时面临反弹，持股过节。×

子论点2 - 有关：中期趋势：上证指数在目前刚跌破60日线，还在宽幅震荡中，总成交已经减少勉强维持在万亿上方，一旦跌破后指数再回到3731的上方，时间就会比较长了。×

论据1.2 - 支持子论点1：昨指数低开震荡下行。盘面上看，燃气、石油、煤炭、电力、钢铁、化工、有色等周期板块遭重挫，个股掀跌停潮；×

子论点2 - 无关：好品种，好赛道，更需要好价格 ×

子论点3 - 无关：买的好而不是买好的 ×

子论点3 - 无关：消费、医疗、科技虽好，莫要贪杯好了 ×

Figure 9: Example 4