Neural Mining of Persian Short Argumentative Texts

Mohammad Yeghaneh Abkenar[1,2], Manfred Stede[1]
[1]University of Potsdam
Applied Computational Linguistics
[2]Bundesdruckerei Gruppe GmbH Berlin
yeghanehabkenar@uni-potsdam.de, stede@uni-potsdam.de

Abstract

Argumentation mining (AM) is concerned with extracting arguments from texts and classifying the elements (e.g., claim and premise) and relations between them, as well as creating an argumentative structure. A significant hurdle to research in this area for the Persian language is the lack of annotated corpora. This paper introduces the first argument-annotated corpus in Persian and thereby the possibility of expanding argumentation mining to this language. The starting point is the English argumentative microtext corpus part 1 (AMT) (Peldszus and Stede, 2015), and we built the Persian variant by machine translation and careful post-editing of the output. We call this corpus *Persian argumentative microtext* (PAMT). Moreover, we present the first results for Argumentative Discourse Unit (ADU) classification for Persian, which is considered to be one of the main subtasks of argumentation mining. We determine the ADUs and their types (claim vs. premise) by two methods: (i) span categorization using the deep learning model of spaCy Version 3.0 (a CNN model on top of Bloom embedding with attention), and (ii) a neural sequence tagger. The results that we obtain with the second approach are comparable to those achieved on the same subtask in AMT and its other translations.

Keywords: Argumentation Mining, Persian Argumentative Corpus, Persian Language Resource

1. Introduction

One of the most essential requirements for developing Natural Language Processing (NLP) solutions for any language is data in that language. Based on the findings of (Paolillo and Das, 2006), out of over 7,000 languages spoken globally, approximately 20 of them have text corpora containing hundreds of millions of words. The language having the most data is by far English, followed by Chinese and Spanish. Japanese as well as Western-European languages are other languages with sizable datasets. The bulk of the languages spoken in Asia and Africa, on the other hand, do not have the training data needed to create reliable, cutting-edge NLP systems. *Low-resource languages* are characterized by being less explored, lacking in resources, being underrepresented in computational tools, and by a lack of annotated data (Singh, 2008). While Persian may not be fully classified as a low-resource language in theoretical terms, according to Joshi et al. (2020), it falls within the category of "The Underdogs" (level 4) in the language race. This designation implies that for Persian there is a significant amount of unlabeled data, but when compared to languages such as English, Spanish, German, Japanese, and French, which belong to level 1, "The Winners," Persian has a smaller amount of annotated data available (Joshi et al., 2020). As noted by Shamsfard (2019), reputable datasets for training and testing Persian systems for important NLP tasks are lacking, although the language is spoken by around 110 million people. Hence, the scarcity of resources and annotated data makes Persian an interesting candidate for research focused on addressing the needs of language with limited resources.

In recent years, progress in the wider field of natural language processing (NLP) such as pre-trained transformer-based models (Devlin et al., 2018) in combination with the increasing availabilty of data of different types has created great potential for almost every area in NLP, including argumentation mining (Stede and Schneider, 2018; Lawrence and Reed, 2020). Argumentation Mining (AM), and specifically the problem of finding argumentation structures in text, has received much attention in the past ten years, but with the research mainly focusing on English.

Broadly, AM can be seen as an extension of sentiment analysis. While sentiment analysis is about "what people think about an entity X", AM extends this to "why people think Y about X", thus uncovering more complex argumentation processes rather than just opinions and sentiments.

Aside from academic interest, AM attracts attention due to its wide range of applications, such as exemplified in the IBM Debater Project.[1] Further, argumentation mining can be used for a variety of important applications such as:

- Decision assistance, using AM in decision making on a controversial issue

[1]https://research.ibm.com/haifa/dept/vst/debater.shtml

2nd EURALI Workshop, pages 30–35

- Product reviews, where AM tools can be applied to product reviews, for instance, to understand why customers value a product.
- Writing support, to assess the quality of argumentative text and provide feedback to authors

Unlike most NLP problems, AM is not a single, straightforward task but a constellation of subtasks. In this paper, we focus on Argumentative Discourse Unit (ADU) classification, which is defined by Hidey et al. (2017, p. 14) as follows:

- *Claim* (Conclusion): A statement articulating the speaker's perspective on a particular issue. It can include predictions, interpretations, evaluations, and expressions of agreement or disagreement with others' assertions.
- *Premise* (Evidence): a statement put forth by the speaker to reinforce a claim, aiming to convince the audience of the claim's validity. While premises can convey opinions, their primary purpose is not to introduce a new viewpoint but rather to support or attack one already expressed by another proposition.

Identifying these components is consistent with the standard definition of an argument, as stated by (Van Eemeren et al., 2004), which requires at least one claim and one statement of evidence, referred to as a premise.

A major contribution of this paper is the free availability of the first annotated Persian corpus for argumentation mining, based on a corpus of short English and German texts introduced by (Peldszus and Stede, 2015). Additionally, we present the first model for argumentation mining for Persian short argumentative texts. Our results on ADU classification can be considered as the first results on this task in Persian. They indicate that sequence tagging models, which have been used for other languages, can also be considered a useful approach for this task in Persian.

2. Related Work

To the best of our knowledge, there are currently no argumentative corpora and results for argumentation mining in Persian, but there is some research on argumentation mining for other non-English languages. Aker and Zhang (2017) created the first annotated Chinese corpus using existing English corpora and manually matched claims and premises with parallel Chinese texts. (Namor et al., 2019) presented an early model for AM for Italian short argumentative texts. By adapting the model created by (Peldszus and Stede, 2015) to Italian and semi-automatically interpreting the original English corpus, they constructed a corpus of Italian microtexts. They utilized two phases for translation: in the first phase, they automatically translated the entire corpus using the DeepL translator service, known for its high-quality translations. In the second phase, they did manual post-editing. They reported results on all four original subtasks of AM according to (Peldszus and Stede, 2015), namely classifying attachment (at), central claim (cc), role (ro) and function(fu). Similarly, (Fishcheva and Kotelnikov, 2019) provided a Russian-language corpus for AM, which is based on machine translation of the Persuasive Essays corpus (Stab et al., 2014) and the AMT corpus. They investigated specifically the subtask of ADU role classification as "proponent" or "opponent".

3. Corpus

3.1. Original Corpus: AMT

The AMT corpus (part 1) consists of 112 short argumentative texts. 22 texts were written by the authors as "proof of concept" of the idea, and 90 texts were collected in a controlled text production experiment in which students wrote short texts, according to suggested length and rhetorical complexity (Peldszus and Stede, 2015).

All texts have been originally written in German and then were professionally (manually) translated into English. Although the texts are short, they are also 'complete', and the argument structure is generally quite clear. The annotation scheme for the AMT corpus has been constructed on the basis of Freeman's approach (Freeman, 2011). Essentially, the ways in which premises and claims are modeled corresponds to a hypothetical dialectical exchange between a proponent and an opponent. We show an example of an annotated text from the AMT corpus in Figure 1.

In the IAA study reported by Peldszus and Stede (2015), three annotators agreed on the complete task (in accordance with the annotation guidelines) with a Fleiss k=0.83 score, and with significantly larger agreement on the fundamental difference between support and attack relations.

The original AMT corpus comes in an XML format. We have extracted texts and their labels using regular expressions (regex) and other extraction packages such as beautifulsoup.[2] Overall, this first part of the AMT contains 112 claims (one for each text), and 464 premises.[3]

[2]BeautifulSoup

[3]Both parts of the English corpus, as well as annotation guidelines and further information, can be found here: argmicro

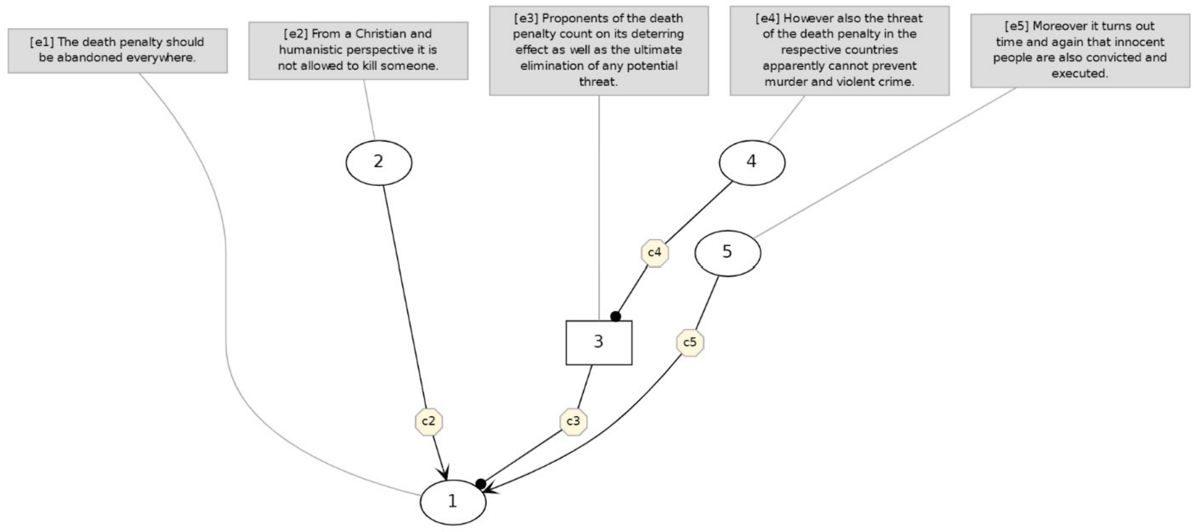

Figure 1: Example text from AMT corpus part 1 (argument 26) and its argumentation structure: Text segmented into ADUs; proponent and opponent role nodes (ellipses versus rectangles); supporting and attacking relations (arrow head, circle head). The first ADU (e1) is the claim, the others are premises.

3.2. Persian Corpus: PAMT

We created PAMT by translating the English AMT (part 1) and mapping all layers of annotations from AMT to PAMT.

Translation. This process was divided into two steps, automatic translation and manual post-editing. For translation, we used the Google translate application programming interface (API).[4] Then, translations were carefully proofread using an XML editor and a customised version of the annotation tool Prodigy,[5] which is a scriptable annotation tool of spaCy.[6]

A maxim for post-editing was to keep the original sentence order and structures of the English texts as parallel as possible to the Persian, so that mapping sentence- and clause-level annotations will be facilitated. English names (such as names of streets, countries, etc.) were translated to Persian. The post-editing was done by an English translator who is an expert in English literature and fluent in Persian.

In the resulting PAMT, the majority of texts consist of four, five, or six segments (ADUs), with an average of 5.1 segments. On average, each text has 3.7 sentences, with an average of 89.5 tokens per text. All other statistics are consistent with those reported in the original paper (Peldszus and Stede, 2015).

[4] https://pypi.org/project/googletrans/
[5] https://prodi.gy/
[6] https://spacy.io/

Annotation. While in our current work, we focus on classifying only the ADU types (claim, premise), we also mapped the relation annotations (support and two types of attack) from AMT to PAMT. Thus, the Persian corpus provides the same tree structures as those that are illustrated in Figure 1.

The annotated corpus and accompanying code is freely available.[7]

4. Experiments and Results

Since AMT and PAMT texts are short, they do not contain non-argumentative material. Therefore, ADU annotation covers the texts completely, so that the task of identifying claims and premises reduced to a binary classification. (This is in contrast to longer texts such as those in the Persuasive Essays corpus by Stab and Gurevych (2014), which can contain non-argumentative sentences.)

We experiment with two separate approaches, span categorization and neural sequence tagging. For the first approach, we divided the corpus into 90 texts for training and 22 texts for evaluation. For the second approach, we used 3-fold cross-validation. In order to prepare the corpus for the classification tasks, we used spacy and hazm[8] for tokenization and adding part of speech (POS) labels.

Span Categorization. As our first approach, we view the task as a span categorization problem. We used spaCy, an open-source library for NLP.

[7] https://github.com/myeghaneh/PAMT
[8] https://github.com/roshan-research/hazm

Recent improvements in spaCy Version 3.0 and Prodigy allow us to label spans even when they are potentially overlapping and nested (though this does not occur in our corpus). Specifically, we use spaCy's *SpanCategorizer* with a CNN model on top of Bloom embeddings with attention.

Neural Sequence Tagger. Following the approach of Chernodub et al. (2019) and Abkenar et al. (2021), we implemented a neural sequence tagger with the Flair NLP framework[9] to identify argumentative units and classify them as claim or premises in PAMT. For sequence labeling tasks, the calculated character-based embeddings are passed into a bidirectional long-short-term memory conditional random field. The tagger learns to assign *B*-$\{C|P\}$ and *I*-$\{C|P\}$ tags to tokens, representing the beginning or the "interior" of claim and premise, respectively. We did a few experiments on different Persian Word embeddings, and we chose Persian FastText embeddings trained over crawls as pre-trained language models (fa-crawl) (Akbik et al., 2019).

We trained on-the-fly in every training mini-batch. This means that the embeddings are not stored in memory. The advantage is that this keeps the memory footprint low. A sample output is shown in Figure 2 with colored labels for the two types of ADUs.

Span Categorization	P	R	F1
PREMISE	0.535	0.523	0.529
CLAIM	0.571	0.545	0.558

Table 1: Class-specific results of ADU classification for PAMT by span categorization.

Sequence Tagging	P	R	F1
PREMISE	0.737	0.304	0.410
CLAIM	0.618	0.734	0.670

Table 2: Class-specific results of ADU classification for PAMT by Sequence tagging using 3-fold cross-validation.

Results. Tables 1 and 2 show a comparison of the class-specific results for our best performing models on PAMT by the two approaches. Overall F1 values are given in Table 3: using span categorization we achieve a micro F1-Score 0.55 for claim vs. premise. Applying the neural sequence tagger with Farsi embeddings yields 0.64 micro F1-Score. These results are, to best of our knowledge, the first that have been reported for this ADU classification task on Persian. In Table 3, we also show the corresponding result reported by Abkenar et al. (2021) for the English AMT corpus.

[9] https://github.com/flairnlp/flair

Method	F1
Persian SpanCategorizer	0.550
Persian NeuralSequneceTagger	**0.636**
Engilsh NeuralSequneceTagger	**0.718**

Table 3: Comparison of PAMT model performance (micro F1-Score) for ADU classification (claim vs. premise) to the result on English AMT by Abkenar et al. (2021).

5. Conclusion and Outlook

Based on the English Argumentative Microtext Corpus, we have produced the first Persian argument-annotated corpus and make it available to a general audience. The AMT corpus was systematically translated into Persian using machine translation (Google Translate API), post-processing, and post-editing of the AMT. Additionally, we projected the entire annotation layer of AMT onto PAMT, making it available for further analyses. Second, we investigated the problem of classifying Argumentative Discourse Units (ADUs) into two classes, "Premise" and "Claim", in Persian. The best performance in Persian was achieved by the Neural Sequence Tagger, with a micro F1-score of 0.64. In comparison to results from experiments with the Italian corpus(Namor et al., 2019), the results were somewhat lower, possibly due to the smaller Persian model in Flair or to differences between the languages. The results of the Neural Sequence Tagger were notably better than those of the SpanCategorizer.

For further research, we plan to conduct more experiments by introducing a corpus similar to the Persuasive Essay Corpus (PEC) (Stab and Gurevych, 2014) in Persian, and using both corpora for cross-domain train/test experiments.

6. Ethics and Limitations

Given our restricted resources for conducting independent studies, our focus was exclusively on Persian, without consideration for other languages spoken in Iran, such as Kurdish, Laki, (Ahmadi et al., 2023b) Baluchi, (Kargaran et al., 2023) and Gilaki (Ahmadi et al., 2023a), which are often deemed low-resource languages. We aspire to broaden our research scope to encompass these languages in the future and encourage collaboration with scientists from these language areas interested in similar topics.

Our study was constrained by a relatively small corpus size, but we prioritized translation quality.

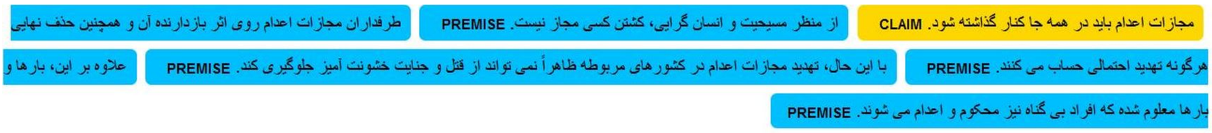

Figure 2: An example text (argument 26) about topic "Introduction of Capital Punishment" in Persian corpus (PAMT) with the prediction of claim from premise by our model.

To address this, we plan to expand the corpus in future versions and incorporate larger datasets. Additionally, our focus solely on ADU classification represents a limitation. Future research will encompass other subtasks within argumentation mining, broadening our findings.

7. Acknowledgement

We would like to thank Mahtab Dadarsefat for her exceptional translation and post-editing contributions, as well as to Sara Shahmohammadi, for her insightful comments on the translation process. Additionally, we thank our anonymous reviewers for the constructive comments, and we extend our heartfelt gratitude to EURALI 2024 for their dedication to promoting diversity and inclusion in language technology by providing resources and tools for lesser-resourced languages.

8. Bibliographical References

Mohammad Yeghaneh Abkenar, Manfred Stede, and Stephan Oepen. 2021. Neural argumentation mining on essays and microtexts with contextualized word embeddings.

Sina Ahmadi, Milind Agarwal, and Antonios Anastasopoulos. 2023a. Pali: A language identification benchmark for perso-arabic scripts. *arXiv preprint arXiv:2304.01322*.

Sina Ahmadi, Zahra Azin, Sara Belelli, and Antonios Anastasopoulos. 2023b. Approaches to corpus creation for low-resource language technology: the case of southern kurdish and laki. *arXiv preprint arXiv:2304.01319*.

Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019. Flair: An easy-to-use framework for state-of-the-art nlp. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics (demonstrations)*, pages 54–59.

Ahmet Aker and Huangpan Zhang. 2017. Projection of argumentative corpora from source to target languages. In *Proceedings of the 4th Workshop on Argument Mining*, pages 67–72.

Artem Chernodub, Oleksiy Oliynyk, Philipp Heidenreich, Alexander Bondarenko, Matthias Hagen, Chris Biemann, and Alexander Panchenko. 2019. Targer: Neural argument mining at your fingertips. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 195–200.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Irina Fishcheva and Evgeny Kotelnikov. 2019. Cross-lingual argumentation mining for russian texts. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 134–144. Springer.

James B Freeman. 2011. *Dialectics and the macrostructure of arguments: A theory of argument structure*, volume 10. Walter de Gruyter.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. 2017. Analyzing the semantic types of claims and premises in an online persuasive forum. In *Proceedings of the 4th Workshop on Argument Mining*. Columbia Univ., New York, NY (United States).

Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. The state and fate of linguistic diversity and inclusion in the nlp world. *arXiv preprint arXiv:2004.09095*.

Amir Hossein Kargaran, Ayyoob Imani, François Yvon, and Hinrich Schütze. 2023. Glotlid: Language identification for low-resource languages. *arXiv preprint arXiv:2310.16248*.

John Lawrence and Chris Reed. 2020. Argument Mining: A Survey. *Computational Linguistics*, 45(4):765–818.

Ivan Namor, Pietro Totis, Samuele Garda, and Manfred Stede. 2019. Mining italian short argumentative texts. In *CLiC-it*.

John C Paolillo and Anupam Das. 2006. Evaluating language statistics: The ethnologue and beyond. *Contract report for UNESCO Institute for Statistics*.

Andreas Peldszus and Manfred Stede. 2015. An annotated corpus of argumentative microtexts. In *Argumentation and Reasoned Action: Proceedings of the 1st European Conference on Argumentation, Lisbon*, volume 2, pages 801–815.

Mehrnoush Shamsfard. 2019. Challenges and opportunities in processing low resource languages: A study on persian. In *International Conference Language Technologies for All (LT4All)*.

Anil Kumar Singh. 2008. Natural language processing for less privileged languages: Where do we come from? where are we going? In *Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages*.

Christian Stab and Iryna Gurevych. 2014. Annotating argument components and relations in persuasive essays. In *Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical papers*, pages 1501–1510.

Christian Stab, Christian Kirschner, Judith Eckle-Kohler, and Iryna Gurevych. 2014. Argumentation mining in persuasive essays and scientific articles from the discourse structure perspective. In *ArgNLP*, pages 21–25.

Manfred Stede and Jodi Schneider. 2018. *Argumentation Mining*, volume 40 of *Synthesis Lectures in Human Language Technology*. Morgan & Claypool.

Frans H Van Eemeren, Robert Grootendorst, and Rob Grootendorst. 2004. *A systematic theory of argumentation: The pragma-dialectical approach*. Cambridge University Press.