Exploring Reasoning Schemes: A Dataset for Syllogism Figure Identification

Shiya Peng[1], Lu Liu[1], Chang Liu[1], Dong Yu[1*]

[1] Beijing Language and Culture University, Beijing, China
pengshiya_blcu@163.com, luliu.nlp@gmail.com, liuchang2014@gmail.com, yudong_blcu@126.com

Abstract. Argument mining aims at extracting structured arguments from texts. Both argument schemes and their structures are studied. A syllogism is a traditional argument scheme and its figures are important in classical logic. Existing research has not yet investigated syllogism figures. Here, we fill this gap by presenting a study on automatic identification of syllogism figures. We prepared a novel dataset of 8.6k syllogisms. We annotated their figures and carried out identification tasks using both supervised and weakly-supervised approaches. Experimental results show that both approaches are adequate.

Keywords: syllogism figure; enthymeme reconstruction; reasoning scheme

1 Introduction

A syllogism is a traditional argument scheme. With the rapid development of argument mining research [1, 2], syllogisms and other types of argument schemes have gained a lot of attention over the last few years [3, 4, 5, 6]. There are now several studies on syllogisms, but the majority have mainly investigated enthymemes [7, 8, 9], i.e., the eclipsed forms of syllogisms. The properties of syllogisms in its complete form is still not much considered.

A syllogism in classical logic can appear in the form of one of its four figures. Figure 1 shows a standard example of the first figure of syllogisms.

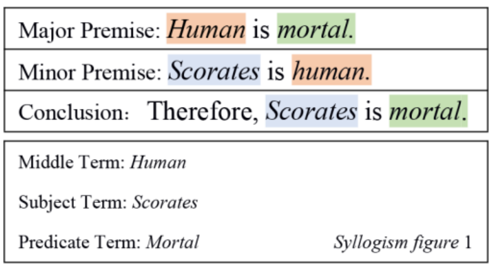

Fig. 1. An example showing the structure of the first figure of syllogism.

*Corresponding author

Previous works in formal logic have suggested that the figures of syllogisms are significant [10, 11]. Grasping the properties of syllogisms, such as figures, can contribute to further understanding of the detailed argument process of syllogisms. As less research focuses on the properties of a specific argument scheme, the syllogism figure is a good cut-off point for further studies on the details of an argument scheme. In this paper, we propose a pilot study on syllogism figures and present a task of syllogism figure identification.

To study syllogism figures, the first issue we have to face is the lack of an available syllogism dataset. Since syllogisms often exist in the form of enthymemes, it is difficult to directly extract syllogisms from natural language texts. An accepted way to get syllogism data is to reconstruct enthymeme [3, 12]. Many works have attempted to investigate enthymeme reconstruction. However, most of them have focused on reconstruction methods based on logic-based frameworks [7, 8], or just tried to reconstruct the eclipsed arguments using natural language on a small scale [9].

In this work, first, we have attempted to carry out large-scale enthymeme reconstruction using natural language to solve the lack of syllogism data for our research. We have constructed a sizable syllogism dataset, comprising more than 8.6k syllogisms with annotated figures. Based on the labeled syllogism dataset, we have then performed the syllogism figure identification task with a supervised approach. Furthermore, we have been interested in whether automatically labeling a large set of unlabeled and similar data could assist deep learning models for the syllogism figure identification task. So we have also attempted a weakly-supervised approach in our investigation.

Our contributions are summed up as follows: A dataset for the research on the syllogism and its figures consisting of more than 8.6k syllogisms is established. An automatic identification task of syllogism figures is presented; and the task performed under both supervised and weakly-supervised approaches is attempted.

2 Related Work

Argument Mining is a multi-disciplinary research field focusing on the study of identifying and analyzing arguments in natural language [13, 14, 2]. Existing works in argument mining have mostly concentrated on extracting arguments from texts in certain domains, such as online debates, online product reviews, philosophical essays, and Wikipedia articles [15, 16, 17].

Several researchers have also attempted to study argument schemes and structures. [4] has studied the task of classifying arguments by analyzing the structure of different inference schemes, such as exemplification or causality. [6] has studied argument schemes in the "results" and "discussion" sections in biological articles. And [5] has proposed a set of guidelines for the annotation of argument schemes with different structures, aiming to overcome the challenges in annotating a broad range of schemes. Although there are several kinds of scheme studies, the syllogism as a traditional argument scheme is still not much considered.

A syllogism is a kind of logical argumentation method in which a conclusion is derived from two premises. The classical structure of a syllogism is determined by its figure. Several researchers have claimed that the figures of the syllogism have a historical importance that can hardly be over-estimated. To build a syllogism dataset for research, the first issue we have to face is enthymeme reconstruction. Many works have provided shared common knowledge as a solution to the reconstruction. Several studies have attempted to reconstruct enthymeme using logic-based frameworks or natural language techniques. But large-scale enthymeme reconstruction is still lacking. So, we attempt to perform large-scale enthymeme reconstruction in this work.

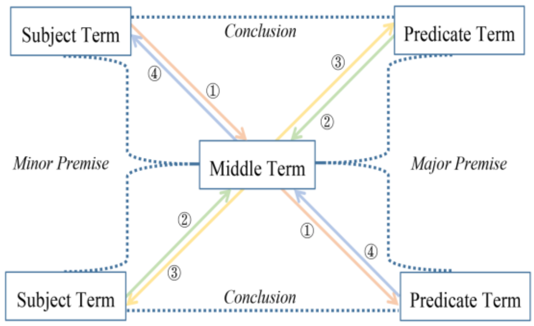

Fig. 2. The relation between the syllogism, terms, and figures. Each dashed line shows the relations between the syllogism and the terms. For example, the minor premise consists of a subject term and a middle term. And each pair of arrows shows the structure of a syllogism figure. For example, cardinal 4 indicates syllogism figure 4, the structural orders of which are from predicate to middle term in the major premise and from middle term to subject term in the minor premise.

Furthermore, [18] has used a small annotated dataset and weakly supervised methods to automatically annotate a large dataset. In this paper, we attempt to investigate the task of identifying syllogism figures using supervised and weakly-supervised methods.

3 The Syllogism and its figure

The syllogism is a classical argument scheme. A typical syllogism includes a minor premise, a major premise, and a conclusion. It has three terms and four figures. Relations between the syllogism, terms and figures are shown in Figure 2.

3.1 Relations between syllogism and terms

A syllogism generally has three terms: the predicate term (P), the middle term (M), and the subject term (S). Each component of the syllogism is an assertion about the relationship between the terms. Traditionally, the major premise depicts the relation between the M and P. The minor premise depicts the relation between S and M. And the conclusion, using deductive reasoning, comes up with a relation between S and P derived from the information provided by the two premises.

3.2 Relations between terms and figures

The syllogisms also has four figures that reflect the structure of reasoning. The positions of the middle term in the major and minor premises define the figure of a syllogism. In figure 1, the orders of the terms in the major and minor premises, respectively, are M-P, S-M; in figure 2, the orders are P-M, S-M; in figure 3, the orders are M-P, M-S; in figure 4, the orders are P-M, M-S.

4 Dataset Creation

An adequate dataset is required for a research on syllogism figures. To our knowledge, few syllogism datasets are available. In this work, we build a syllogism dataset for our research. The approach we used to build our dataset is based on enthymeme reconstruction. An enthymeme is a syllogism that has one of its premises omitted, and it can be restored to a standard syllogism if the missing premise is added back. SNLI [19] includes some inference sentence pairs that satisfy the definition of enthymeme, and the difficulty of restoring the missing premise of these enthymemes is indicated at an appropriate level. Therefore, we chose the entailment part of SNLI as the source data for our annotation.

Table 1. Data statistics of each syllogism figure.

Category	Number	Proportion
figure 1	6428	74.10%
figure 2	1808	20.80%
figure 3	363	4.20%
figure 4	74	0.85%

Our dataset was annotated by ten annotators. All of them are at the graduate level with the linguistics background. This ensures that they have the basic ability to annotate the data for our reconstruction task. To guarantee the annotation quality, the annotators were required to attend a pre-training procedure. We utilized Fleiss' kappa [20] to compute IAA between the annotators. After the training, the Kappa score was raised from 0.469 to 0.746, indicating substantial agreement (within [0.6, 0.8]). The formal annotation included three steps: analyzing the enthymeme, restoring the missing premise, annotating the syllogism figure.

Our data source is the entailment part of SNLI. The inference relations between entailment pairs in the dataset are mostly repetition or an abbreviation. They are not enthymemes. These invalid sentence pairs should be filtered out. The annotators were required to check the entailment pairs to determine whether they were enthymemes in the first step. Then the valid labeled enthymemes would be further analyzed by the annotator. They would decide which one of the premises was missing. Then they would be required to restore the missing premise using the terms of the syllogism. After completing the syllogism, the annotators would annotate the figure of the syllogism. In the annotation process of step 1, totally we annotated 79,700 data, and finally got 8,673 valid annotated enthymemes. The statistics of the four syllogism figure types are shown in Table 1.

5 Experiments

We address the syllogism identification task as multi-class classification. Experiments were performed using two different approaches: supervised and weakly-

Table 2. Performance of different models using two approaches. sp stands for supervised method, wsp stands for weakly-supervised method.

	maP	miP	maF1	miF1	ACC
sp_SVM	0.55	0.74	0.29	0.74	0.74
sp_BiLSTM	0.59	0.85	0.53	0.85	0.86
sp_BERT	0.62	0.91	0.61	0.91	0.92
wsp_BERT	0.62	0.90	0.61	0.90	0.90

supervised approaches. The widely used Macro-Average and Micro-Average scores were used as evaluation methods.

5.1 Supervised Approach

Only human-annotated syllogisms data was used in this approach. Our labeled dataset was split into training (6937 components), validation (868 components) and test (868 components) sets, keeping a component proportion similar to that in the dataset. We chose three methods for the supervised approach, including Support Vector Machines (SVM) [21], Bidirectional Long Short-Term (BiLSTM) [22] and Bidirectional Encoder Representations from Transformers (BERT) [23]. These methods were used separately in our identification task.

5.2 Weakly-supervised Approach

A large scale of weakly labeled data, annotated by classifiers, was used in this approach. E-snli [24] is an entailment explanation dataset. It contains a certain amount of data which has structures similar to the syllogism, and this makes it a suitable source for the weakly-supervised approach. In this experiment, E-snli was selected as our unlabeled data source. BERT was selected as the empirical method in this approach. We first used our labeled data to train a classifier to identify the similar syllogism data in the dataset. Then the human-annotated data were randomly divided into three training sets that were used to train another classifier to get the data with labeled figures for the weakly-supervised approach. These weakly labeled data were used to train BERT for our syllogism figure identification task.

5.3 Results and Analysis.

Table 2 contains the experimental results of the supervised and weakly supervised approaches. The accuracies of the models using the supervised approach are 0.74, 0.82 and 0.92. BERT has performed better than the other two methods on the task. This may be because the BERT model, which has been pre-trained on large-scale data of common sense, has a lot of external knowledge, and was able to better enhance the ability to perform the task on a small dataset. The best model achieves 0.92 in accuracy using the supervised approach and 0.90 in accuracy using the weakly-supervised approach. The performance of the model using the weakly supervised approach and trained with large data is same to the model using the supervised approach.

We conjure that both of the approaches have the ability to identify syllogism figures. Among the metrics for the BERT model using a supervised approach, maP is 0.62, and miP is 0.91. The Macro-Average metrics are much lower than the Micro-average metrics. They have a significant difference. This difference indicates that although the average accuracy over the four classes is good, a class with a small size may not be identified effectively. This is because the dataset has an imbalanced class distribution. For this issue, we will attempt other learning methods and focus on the way of how to better improve the model’s learning of labeling with a small dataset in the future.

6 Conclusion

This work has investigated a task of syllogism figure identification. We have prepared a syllogism dataset with annotated figures and carried out the task using both supervised and weakly-supervised approaches. Our results have shown that using a weakly-supervised approach is as good as using a supervised approach for the identification task. In future work, we propose to utilize these automatic approaches to assist human annotation for syllogism. And more balanced data will also be used in future syllogism dataset construction.

Acknowledgments. This work is funded by the Humanity and Social Science Youth foundation of Ministry of Education (19YJCZH230) and the Fundamental Research Funds for the Central Universities in BLCU (No.17PT05).

References

1. Manfred Stede and Jodi Schneider. 2018. Argumentation mining. Synthesis Lectures on Human Language Technologies, 11:1–191
2. John Lawrence and Chris Reed. 2019. Argument mining: A survey. Computational Linguistics, pages 1–55.
3. Douglas Walton, Chris Reed, and Fabrizio Macagno. 2008. Argumentation Schemes. Cambridge University Press.
4. Vanessa Wei Feng and Graeme Hirst. 2011. Classifying arguments by scheme. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pages 987–996, Portland, Oregon, USA. Association for Computational Linguistics.
5. Elena Musi, Debanjan Ghosh, and Smaranda Mure-san. 2016. Towards feasible guidelines for the annotation of argument schemes. In Proceedings of the Third Workshop on Argument Mining, pages 82–93. Association for Computational Linguistics.
6. Nancy Green. 2018. Proposed method for annotation of scientific arguments in terms of semantic relations and argument schemes. In Proceedings of the 5th Workshop on Argument Mining, pages 105–110, Brussels, Belgium. Association for Computational Linguistics.
7. Elizabeth Black and Anthony Hunter. 2012. A relevance-theoretic framework for constructing and deconstructing enthymemes. Journal of Logic and Computation, 22:55–78.
8. Pavithra Rajendran, Danushka Bollegala, and Simon Parsons. 2016. Contextual stance classification of opinions: A step towards enthymeme reconstruction in online reviews. In Proceedings of the First Workshop on Argumentation Mining, pages 31–39. Association for Computational Linguistics
9. Olesya Razuvayevskaya and Simone Teufel. 2017. Finding enthymemes in real-world texts: A feasibility study. Argument & Computation, 8:1–17.
10. Hubert Hubien. 1975. John buridan on the fourth figure of the syllogism. Revue Internationale de Philoso-phie, 113(113):271–285
11. Louis Dickstein. 1978. The effect of figure on syllogistic reasoning. Memory & Cognition, 6:76–83.
12. S.A.Hosseini, S.Modgil, and Odinaldo Rodrigues. 2014. Enthymeme construction in dialogues using shared knowledge. Frontiers in Artificial Intelligence and Applications, 266:325–332
13. Marco Lippi and Paolo Torroni. 2016. Argumentation mining: State of the art and emerging trends.ACMTrans. Internet Technol., 16(2):10:1–10:25.
14. Elena Cabrio and Serena Villata. 2018. Five years of argument mining: a data-driven analysis. IJCAI-18, pages 5427–5433. International Joint Conferences on Artificial Intelligence Organization.
15. John Lawrence, Chris Reed, Colin Allen, Simon McAlister, and Andrew Ravenscroft. 2014. Mining arguments from 19th century philosophical texts using topic based modeling. In Proceedings of the First Workshop on Argumentation Mining, pages 79–87, Baltimore, Maryland. Association for Computational Linguistics.
16. Ehud Aharoni, Anatoly Polnarov, Tamar Lavee, Daniel Hershcovich, Ran Levy, Ruty Rinott, Dan Gutfreund, and Noam Slonim. 2014. A benchmark dataset for automatic detection of claims and evidence in the context of controversial topics. In Proceedings of the First Workshop on Argumentation Mining, pages 64–68.
17. Mara Garca-Villalba and Patrick Saint-Dizier. 2012. A framework to extract arguments in opinion texts. International Journal of Cognitive Informatics and Natural Intelligence, 62–87
18. Pavithra Rajendran, Danushka Bollegala, and Simon Parsons. 2018. Is something better than nothing?Automatically predicting stance-based arguments using deep learning and

small labelled dataset. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, pages 28–34
19. Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pages 632–642.
20. Joseph Fleiss. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin, 76:378
21. Corinna Cortes and Vladimir Vapnik. 1995. Support vector network. Machine Learning, 20:273–297.
22. Alex Graves and Jrgen Schmidhuber. 2005. Framewise phoneme classification with bidirectional lstm andother neural network architectures. Neural networks: the official journal of the International Neural Net-work Society, 18:602–10
23. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: pre-training of deep bidirectional transformers for language understanding. CoRR, abs/1810.04805.
24. Camburu, Oana-Maria & Rocktäschel, Tim & Lukasiewicz, Thomas & Blunsom, Phil. (2018). e-SNLI: Natural Language Inference with Natural Language Explanations. In Advances in Neural Information Processing Systems 31, pages 9539–9549