Improving Evidence Detection using Warrants as External Knowledge

Keshav Singh[†] Naoya Inoue[†,‡] Paul Reisert[‡] Kentaro Inui[†,‡]

[†]Tohoku University [‡]RIKEN Center for Advanced Intelligence Project (AIP)
{keshav.singh29, naoya-i, inui}@ecei.tohoku.ac.jp
paul.reisert@riken.jp

1 Introduction

Evidence detection [12] is an important sub-task of argument mining and finds an application in a wide range of natural language comprehension tasks like identifying support and attack relation [3], essay evaluation [19] and identifying entailment between sentences[2]. Given a claim and a set of candidate premises (or evidences), the evidence detection task identifies the correct link between the given claim and a candidate premise. Suppose the following example:

(1) Claim: *Mary is probably pregnant.*
Candidate Premise 1: *The pregnancy test was positive.*
Candidate Premise 2: *Non-taxable items, promotes illegal activity*

The task is to identify Candidate Premise 1 as "evidence" and Candidate Premise 2 as "non-evidence". For identifying the link between a claim and premise, a warrant (e.g. "*Positive test usually indicates one is pregnant*") plays an important role. In the past literature related to argumentation, the use of warrants as a way to establish the link between a claim and premise has been studied extensively [7, 6, 8]. [14, 7] explain the role of warrants as an external evidence, useful in linking the premise and the claim. The Toulmin model [20] is also widely accepted to be both well-structured and general, and has been shown to be useful for identifying relations between argumentative components in many argumentative texts. Nevertheless, previous work on evidence detection [12] and argumentative relation identification [3, 13, etc.] does not use warrants as a hint for prediction. Therefore, it remains an open-ended question as to whether or not warrants can be used as external knowledge for improving evidence detection.

This paper explores a new approach for identifying the link between a claim and premise by using warrants as external knowledge useful for evidence detection. We hypothesize that a set of warrants related to the claim functions as a hint for claim-premise link identification. We model a deep architecture that captures the link between a premise and claim to explore the applicability of external knowledge in form of warrants for evidence detection. Our preliminary experiments demonstrate the effectiveness of using warrants for the task of evidence detection.

2 Related work

Evidence detection has received wide attention over the years. [12] devised a method for detecting relevant claims to a given debate topic. [17] created a benchmark dataset for identifying context-dependent evidence. However, little work has been done for integrating warrants as background knowledge for detecting evidence.

The work similar to ours comes under the domain of argument mining and argument reasoning, including both supervised and unsupervised models to identify relations between sentence pairs. Previous argumentative relation identification approaches relied more on features extracted from argument component, e.g., semantic similarity, word pairs, textual entailment and so on. However incorporating external knowledge to identify such relations is still an untouched segment. [13]explore discourse structure features for argumentative relation identification. Their work focuses on context aware argumentative relation mining that uses features extracted from writing topics as well as from adjacent context sentences. Similar to this [11] focuses on macro-level information(e.g. argumentative flow) constructed using argumentative relations in a document. [19] approach the task of relation identification by identifying argumentative discourse structures and consider discourse markers as misleading or insufficient to identify argumentative relations, thus introduce their own feature sets. Also instead of taking sentences as inputs their inputs vary from clauses to multiple sentences. [15] present first data-driven model of argumentation structure where they optimize argumenta-

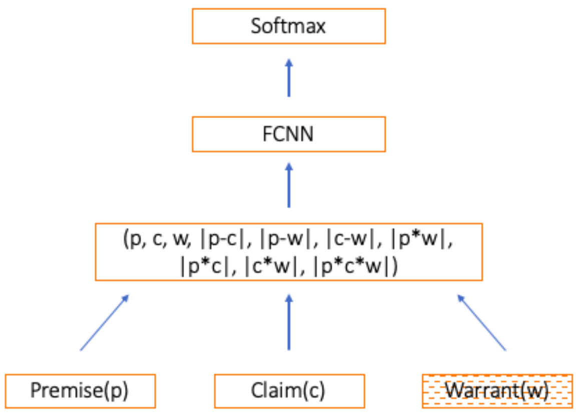

Figure 1: Model Architecture

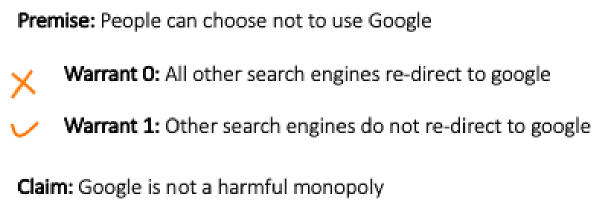

Figure 2: Instance from Argument Reasoning Comprehension task.

tion structure globally for the complete sequence of input components. But they do not account for non-relevant components appearing in the argumentative text. All the above approaches work towards identifying either (*support*/*attack*) or (*support*, *attack or none*) relation between argumentative text.

In contrast to the work mentioned above, the approach in this paper focuses on identifying correct premise(evidence) from a set of evidence in a supervised way. For this reason we use warrants to link the correct pair of claim and evidence. Where warrants are established to link the support of source towards target. In addition, the approach tries to incorporate background knowledge in the form of warrants which we believe is the first attempt towards this task. Our idea, correlates to the approach made by [1], where they try to fill in the gap between claim and premise by letting annotators write down implicit warrants. They show that using manually-compiled premises improves similarity-based claim matching and that premises generalize to unseen user claims. But they concluded only with a preliminary analysis due to large variance in the responses. Also [3] model a deep learning architecture with identifying relations of attack and support between natural language arguments in text, by classifying pairs of pieces of text as attack, support or neither attack nor support relations. We combine the above two approaches partially and use warrants as external knowledge so as to better identify the correct premise as evidence for the respective claim using warrants.

3 Model

3.1 Key idea

We assume that warrants are like external knowledge fragments necessary to link the claim and premise better together. We propose a warrant aware model to better understand relations between C and its P with the use of these knowledge fragments. Also, since these warrants are not available exclusively, we assume that given a claim, we propose to link together warrants relative to that particular claim. The reason we do this is, since warrant are rule-like general statements [21] therefore its applicability is justified over a wide variety of premises given a claim. And in such way we can better identify the relations even when given adversarial instances.

3.2 Architecture

We extend Conneau [5]'s architecture, which is originally proposed for classifying a pair of sentences. Figure 1 depicts our architecture used for binary classification of supporting evidence and non-supporting evidence detection task $L = \{\mathsf{Support}, \mathsf{Non\text{-}relevant}\}$. Here, (p, c, w) refers to encoded vector representations of $(premise, claim, warrant)$ in B_m. To get encode sentences into a fixed-size vector, we use a bidirectional LSTM [18]. We concatenate the hidden states of forward and backward LSTM which reads the sentence in opposite directions. Before getting the fixed size vector, we apply max pooling [4] by selecting maximum value over each dimension of hidden units.

After generating the vector representation of sentences, 3 matching methods are applied to extract relations between (p, c, w).

1. Concatenation of individual representation (p, c, w)
2. Element-wise product $(p*w, p*c, c*w, c*w*p)$
3. Absolute element-wise difference $(|p-c|, |c-w|, |p-w|)$

All the resulting vectors are concatenated and fed to a softmax classifier, connected by fully connected neural network. The softmax layer predicts the label $\in L$ for the relation between C and P.

Hyperparameter	Value
Dropout	0.5
BiLSTM size	32
Batch size	32
Embedding size	100

Table 1: Hyperparameters used in our experiments.

Model	Accuracy
Baseline model	72.71 ±2.17
+ Warrant	76.74 ±1.53

Table 2: Performance of claim-premise link detection (averaged over 5 runs with different random seeds).

4 Experiments

4.1 Dataset

For our task, we use Argument Reasoning Comprehension Corpus (ARCC), the annotated data provided by [9] for SemEval task 2018 [10]. This dataset fits well in our hypothesis of using warrants for claim and premise evidence detection. An instance of the dataset is shown in Figure2. Although the initial task concerning *ARC* dataset is to identify the correct warrant given two choices as shown in Figure 2. We use this dataset by extracting all the correct choice warrants and then modifying the dataset so as to fit it in our assumption (see Sec 3.1).

The dataset covers various topics such as technology, politics, general issues, and so on. In total, we found that there are 169 different claims and the dataset is very diverse, small and involves multiple domains which makes the task challenging. In total, we have 1,210 instances as training data and 444 instances as test data. We slightly change this dataset and collect all correct warrants.

After processing we are left with each instance as the tuple T of type (P, C, W_c), where W_c is the correct warrant for the given instance. We later use this data for further use.

In accordance with our assumptions, we make a new dataset from the given (P, C, W_c) instances. Given the tuples of (*Premise*,*Claim* and *Correct warrant*), we link together all the W_c that are associated to their respective claims.

$$C_i \longrightarrow [w_1, w_2, w_3...]_i \ , \ i \in (1, 169)$$

Next, we randomly choose one warrant (W_{ci}) for C_i from its respective set $[w_1, w_2, w_3...]_i$ and make a new dataset along with (P). Hence, creating the new dataset (P, C_i, W_{ci}), we move to the next step of generating negative instances with randomly choosing premise for the (C_i, W_{ci}) tuple. We treat the initial 1,210 instances (P, C_i, W_{ci}) as Positive Dataset P_D and label them $\{1\}$. Next, we randomly select P_r for P_r in (P_r, C_i, W_{ci}) making sure that P_r does not match with its original C as in P_D. This approach makes sure that P_r and C_i forms neither support nor attack relation between them. We call these tuples as Negative instances (P_r, C_i, W_{ci}) and term this dataset as Negative Dataset N_D and label them $\{0\}$. Further we combine these datasets and get 2420 instances of {*Support, Non-relevant*} relation, labeled as {*1,0*} respectively. Later we apply the same approach to testing dataset and get 888 testing instances. We term this new combined dataset as Benchmark $B_m = P_D \in (P, C_i, W_{ci}, 1) \cup N_D \in (P_r, C_i, W_{ci}, 0)$.

4.2 Settings

Word embeddings are initialized with 100-dimensional GloVe vectors [16]. The words that do not appear in GloVe are initialized with a random value. We use the same hyperparameters for training both models. The values of hyperparameters used in our experiments are summarized in Table 1. We employ early stopping with a patience of 15. Adagrad [22] is used for optimization.

During training, we divide the dataset of ARCC into a training dataset and validation dataset with 8:2 split. We then get the best weights on the validation dataset and use it in the testing phase. To make sure that the dataset is evenly distributed in both the training and validation set, we used a stratified shuffle split. For the testing phase, we use 888 instances from the test dataset of ARCC.

In our experiments, we compare two models: (i) *baseline model*: our model without the warrant encoding layer, and (ii) *proposed model*: our model trained on B_m dataset along with warrants.

4.3 Results

The results of our experiments are shown in Table 2. We observe that using warrants as background knowledge significantly improves the performance of our model. We attribute this to the fact that a premise and a claim are more likely to be related (i.e., the premise acts as evidence for the claim) if a warrant links both of them.

5 Conclusion and Future work

This paper has presented a first approach to incorporate background knowledge in the form of warrants to identify the correct premise from a given set, for a given claim. The results reveal that having in-domain knowledge in small fragments can help identify the correct premise significantly. Although the results look promising but there is still room for further improvement for exploring this hypothesis.

The results obtained motivate us to dwell deeper in the direction of using external knowledge. For future work, we plan to investigate graphical way of representing domain specific knowledge and incorporate it to improve evidence detection task. Also, we will extend our approach to different corpora with graphical knowledge representation.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 15H01702 and JST CREST Grant Number JPMJCR1513, including AIP challenge.

References

[1] Filip Boltuzic and Jan Šnajder. Fill the gap! analyzing implicit premises between claims from online debates. In *Proceedings of the Third Workshop on Argument Mining (ArgMining2016)*, pages 124–133, 2016.

[2] Samuel R Bowman, Gabor Angeli, Christopher Potts, and Christopher D Manning. A large annotated corpus for learning natural language inference. *arXiv preprint arXiv:1508.05326*, 2015.

[3] Oana Cocarascu and Francesca Toni. Identifying attack and support argumentative relations using deep learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1374–1379, 2017.

[4] Ronan Collobert and Jason Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning*, pages 160–167. ACM, 2008.

[5] Alexis Conneau, Douwe Kiela, Holger Schwenk, Loic Barrault, and Antoine Bordes. Supervised learning of universal sentence representations from natural language inference data. *arXiv preprint arXiv:1705.02364*, 2017.

[6] James B Freeman. Relevance, warrants, backing, inductive support. *Argumentation*, 6(2):219–275, 1992.

[7] James B Freeman. Argument strength, the toulmin model, and ampliative probability. *Informal Logic*, 26(1):25–40, 2008.

[8] G Thomas Goodnight. Legitimation inferences: An additional component for the toulmin model. *Informal Logic*, 15(1), 1993.

[9] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. The argument reasoning comprehension task: Identification and reconstruction of implicit warrants. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, volume 1, pages 1930–1940, 2018.

[10] Ivan Habernal, Henning Wachsmuth, Iryna Gurevych, and Benno Stein. Semeval-2018 task 12: The argument reasoning comprehension task. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 763–772, 2018.

[11] Tatsuki Kuribayashi, Paul Reisert, Naoya Inoue, and Kentaro Inui. Towards exploiting argumentative context for argumentative relation identification. 2018.

[12] Marco Lippi and Paolo Torroni. Context-independent claim detection for argument mining. In *IJCAI*, volume 15, pages 185–191, 2015.

[13] Huy Nguyen and Diane Litman. Context-aware argumentative relation mining. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1127–1137, 2016.

[14] E Michael Nussbaum. Argumentation, dialogue theory, and probability modeling: Alternative frameworks for argumentation research in education. *Educational Psychologist*, 46(2):84–106, 2011.

[15] Andreas Peldszus and Manfred Stede. Joint prediction in mst-style discourse parsing for argumentation mining. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 938–948, 2015.

[16] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[17] Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M Khapra, Ehud Aharoni, and Noam Slonim. Show me your evidence-an automatic method for context dependent evidence detection. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 440–450, 2015.

[18] Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.

[19] Christian Stab and Iryna Gurevych. Identifying argumentative discourse structures in persuasive essays. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 46–56, 2014.

[20] Stephen E Toulmin. *The uses of argument*. Cambridge university press, 2003.

[21] Bart Verheij. Evaluating arguments based on toulmins scheme. *Argumentation*, 19(3):347–371, 2005.

[22] Matthew D Zeiler. Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.