Recognizing Textual Entailment Using Probabilistic Inference

Lei Sha, Sujian Li, Tingsong Jiang, Baobao Chang, Zhifang Sui
Key Laboratory of Computational Linguistics, Ministry of Education
School of Electronics Engineering and Computer Science, Peking University
Collaborative Innovation Center for Language Ability, Xuzhou 221009 China
{shalei, lisujian, tingsong, chbb, szf}@pku.edu.cn

Abstract

Recognizing Text Entailment (RTE) plays an important role in NLP applications including question answering, information retrieval, etc. In recent work, some research explore "deep" expressions such as discourse commitments or strict logic for representing the text. However, these expressions suffer from the limitation of inference inconvenience or translation loss. To overcome the limitations, in this paper, we propose to use the predicate-argument structures to represent the discourse commitments extracted from text. At the same time, with the help of the YAGO knowledge, we borrow the distant supervision technique to mine the implicit facts from the text. We also construct a probabilistic network for all the facts and conduct inference to judge the confidence of each fact for RTE. The experimental results show that our proposed method achieves a competitive result compared to the previous work.

1 Introduction

For the natural language, a common phenomenon is that there exist a lot of ways to express the same or similar meaning. To discover such different expressions, the Recognising Textual Entailment (RTE) task is proposed to judge whether the meaning of one text (denoted as H) can be inferred (entailed) from the other one (T)(Dagan et al., 2006). For many natural language processing applications like question answering, information retrieval which involve the diversity of natural language, recognising textual entailments is a critical step.

PASCAL Recognizing Textual Entailment (RTE) Challenges (Dagan et al., 2006) have witnessed a variety of excellent systems which intend to recognize the textual entailment instances. These systems mainly employ "shallow" techniques, including heuristics, term overlap, syntactic dependencies(Vanderwende et al., 2006; Jijkoun and de Rijke, 2005; Malakasiotis and Androutsopoulos, 2007; Haghighi et al., 2005). As Hickl (2008) stated, the shallow approaches do not work well for long sentences for the missing of underlying information which needs to be mined from the surface level expression.

Recently, some deep techniques are developed to mine the facts latent in the text. Hickl (2008) proposed the concept of discourse commitments which can be seen as the set of propositions inferred from the text, and used a series of syntax-level and semantic-level rules to extract the commitments from the T-H pairs. Then the RTE task is reduced to the identification of the commitments from T which are most likely to support the inference of the commitments from H. From the work of Hickl (2008), we can see that a deep understanding of text is critical to the RTE performance and discourse commitments can serve a good media to understanding text. However, the limitation of Hickl (2008)'s work is, the extracted discourse commitments are still from the original text and do not explore the implicit meaning latent behind the text.

Another kind of deep methods involves first transferring natural language to logic representation and then conducting strict logic inference based on the logic representations (de Salvo Braz et al., 2006; Tatu and Moldovan, 2006; Wotzlaw and Coote, 2013). Through logic inference, some implicit knowledge behind the text can be mined. However, it is not easy to translate the natural language text into formal logic expressions and the translation process inevitably suffer from great information loss.

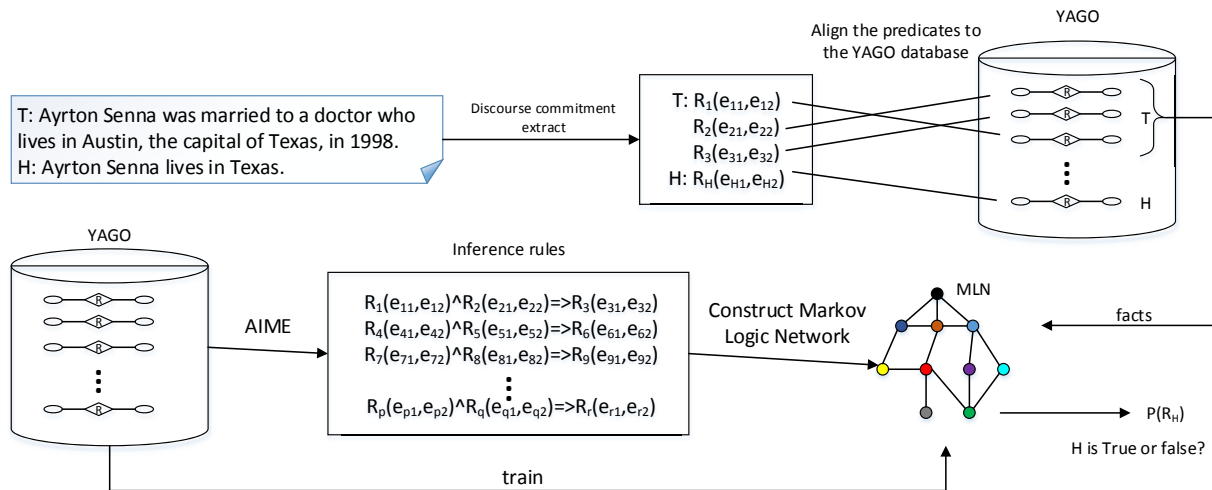Through analysis above, in our work, we pro-

1620

Figure 1: The Framework of our RTE system

pose to use the predicate-argument structure to represent the extracted discourse commitments. Inspired by the work of (Mintz et al., 2009), we make use of the external knowledge YAGO and borrow the distant supervision technique to mine implicit facts for the extracted predicates. For example,

Ayrton Senna was married to a doctor who lives in Austin, the capital of Texas, in 1998.

We translate this example into the predicate-argument structures such as bemarried(Senna, doctor), livein(doctor, Austin), captial(Austin, Texas). Then through distant supervision, we can get some new facts livein(Senna, Austin), livein(Senna, Texas).

To judge the confidence of the new facts, we construct a probabilistic network with all the facts and adopt the Markov Logic Network (MLN) to calculate the probability of each new fact, which can be further used to recognize text entailments.

2 Our RTE System

To make full use of the underlying information in sentences and lessen the effect brought by natural language's vagueness, we design a RTE system which is composed of three stages, as shown in Figure 1.

First, we decompose the sentences in T-H pair to a series of discourse commitments as Hickl (2008) did. Since the syntax of these commitments are very simple, we can directly transform them to predicates (or 3-arg tuples). Then we use YAGO, a large semantic database including several thousands relations, to provide distant supervision. The predicates in T are matched to YAGO facts due to some metrics. At last, we use the Markov Logic Network(MLN) (Richardson and Domingos, 2006) to infer the correctness of the predicates in H. The MLN is constructed using the inference rules ("soft" rules) generated by AMIE system (Galárraga et al., 2013) on YAGO. Each rule has a weight which should be trained by real world facts. Using this framework, we can apply the "soft" logic to the textual entailment recognition task.

2.1 Extracting Predicates from Sentences

Discourse commitment is our baseline system proposed by Hickl (2008) which can decompose one sentence to a series of shorter and simpler sentences which completely contain the origin sentence's information. One of the advantages of discourse commitments is that it can use a lot of syntax-level and semantic level rules to extract the underlying information of one sentence. For example, the following T-H pair can be decomposed as Figure 2. The discourse commitments of T contains the information: Ayrton Senna married in 1998, which is not easy to be captured by "shallow" methods.

T : Ayrton Senna was married to a doctor who lives in Austin, the capital of Texas, in 1998.

H : Ayrton Senna lives in Texas.

Since we need to infer new facts using the extracted commitments, we transfer all the commit-
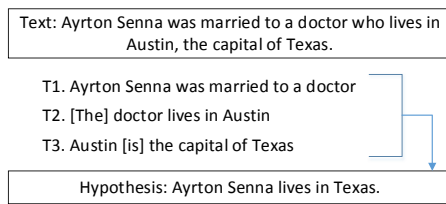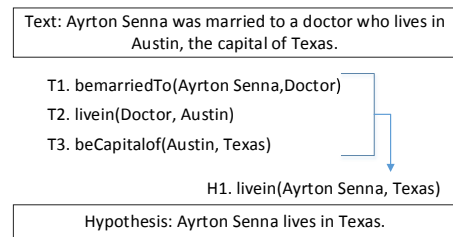
Figure 2: Text Commitments Example

Figure 3: Text Predicates Example

ments to predicates. For example, the commitments in Figure 2 can be transformed to the predicates (or triples) shown in Figure 3. We use RE-VERB (Fader et al., 2011) to extract the triples (predicate + 2 Arguments). To make the inference process in the next section more convenient, we order that all of the arguments should be NPs. Therefore, we check if the arguments in the triples contain or have overlap with any of the NPs, replace it with that NP, and the predicates are successfully extracted.

2.2 Distant supervision with YAGO

The goal of distant supervision is to use the knowledge of YAGO (Mahdisoltani et al., 2014) to help the textual entailment recognition task. The facts in YAGO have various type of connections with each other. We think this connection is very useful for RTE. Therefore, the predicates in the T-H pair should be matched to the YAGO facts for making advantage of the connection information.

Since YAGO is very large, the common predicates can easily matched to YAGO facts in most cases. However, YAGO cannot contain every predicate in T. We run DIRT (Lin and Pantel, 2001) system on 1GB text random sampled from Gigaword corpus and for each predicate we choose the top-10 similar predicates as its synonymous predicate. If the origin predicate cannot be found in YAGO, we instead check for the top-10 similar predicates. If we still cannot find a match, that means this predicate has very little connection with other predicates and cannot be supervised by YAGO.

2.3 Probabilistic Inference

The goal of MLN (Richardson and Domingos, 2006) is to implement the probabilistic inference, or "soft" logic inference. MLN is constructed by an inference rule base. Each rule has a weight which needs to be well trained by real word facts. We use AMIE to mine inference rules from YAGO.

AMIE[1] (Galárraga et al., 2013) is a state-of-the-art inference rule mining system. The motivation of AMIE is that KBs themselves often already contain enough information to derive and add new facts. If, for example, a KB contains the fact that a child has a mother, then the mother's husband is most likely the father.

$$motherOf(m, c) \wedge marriedTo(m, f) \Rightarrow fatherOf(f, c)$$

AMIE can mine such inference rules from large KBs. The inference rules can be directly used for constructing Markov logic network in the next section. In addition, the process of mining inference rules is quite efficient so that it is very helpful for our RTE task.

We use AMIE only to extract the inference rules. After the inference rules are prepared, we can construct a MLN. We give the related facts in YAGO to the MLN, then the weights of each inference rule can tune to a best fit for these facts. After the weights of each inference rule are well trained, the MLN is well prepared to use. Given the predicates in T, we first select all the related rules to construct a simple MLN, and then give the MLN some facts. After that, the MLN will calculate the probabilities of the unknown new facts. The arguments of the new facts are the permutations of all the ground atoms (or entities). For example, if we give the facts "hasChild (Cliton, Chelsea)" and "IsMarriedto (Cliton, Hillary)", the MLN will output the probability of "hasChild (Cliton, Hillary)", "hasChild (Hillary, Chelsea)", "IsMarriedto (Cliton, Chelsea)", etc. Obviously, the probability of "hasChild (Hillary, Chelsea)" may be the highest, so that it is most likely to be true. The MLN constructing and inferring can be implemented by

---

[1]http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/amie/

Approach	Accuracy
Term overlap (Zanzotto and Moschitti, 2006)	67.50%
Graph Matching (MacCartney et al., 2006)	65.33%
Classification-Based (Hickl et al., 2006)	77.00%
Discourse Commitment (Hickl, 2008)	84.93%
Strict logic (Tatu and Moldovan, 2006)	71.59%
Our Framework	85.16%

Table 1: Performance of Inference based RTE

Alchemy[2], which provides a series of algorithms for statistical relational learning and probabilistic logic inference, based on the Markov logic representation.

After the new facts are inferred, we check if these facts can cover the predicates in H. If so, we decide T entails H.

3 Experiment

We evaluate the performance of our framework for RTE on the PASCAL RTE-2[3] and RTE-3[4] datasets, which has 1600 examples. We use the YAGO2 for aligning predicates and mining inference rules. YAGO2 contains more than 940K facts and about 470K entities. We run the AMIE system on YAGO2 for only one time to get all inference rules (about more than 1.8K in total). For each T-H pair, we only choose a portion of related inference rules to construct MLN. The chosen rules must contain at least one predicate which occurred in the predicates of T-H pair. We only use the MLN to infer when the discourse commitment paraphrasing cannot identify a T-H pair as "Entailment", which is a back-off method.

We compare our result with 5 baseline systems: (1) Zanzotto and Moschitti (2006)'s simple term-overlap measure, (2) MacCartney et al. (2006)'s semantic graph-mapping approach, (3) Hickl et al. (2006)'s classification-based term alignment approach. (4) Hickl (2008)'s discourse commitment based Alignment, (5) Tatu and Moldovan (2006)'s strict logic based method. The comparison of the 5 baselines and our framework is shown in Table 1. Since we only need to judge "Yes" or "No" for the 1600 examples, the precision is equal to the recall, so that we only report the precision.

According to the Table 1, the performance of our framework is higher than Hickl (2008)'s baseline, which is significant (Wilcoxon signed-rank test, $p < 0.05$). The reason is that we have added the inference portion to Hickl (2008)'s method. Therefore, some T-H pairs which had to be judged by semantic reasoning can be corrected by our framework. For instance, T is "Hughes loved his wife, Gracia, and was absolutely obsessed with his little daughter Elicia." and H is "Gracia's daughter is Elicia." It is not easy for the former baselines to recognize this entailment, but our framework can easily recognize it to be "true". In this way, our framework has achieved a higher result.

4 Related work

Textual Entailment Recognizing (RTE) task has been widely studied by many previous works. Firstly, the method based on similarity and overlap (Malakasiotis and Androutsopoulos, 2007; Jijkoun and de Rijke, 2005; Wan et al., 2006). This kind of methods can help solve the paraphrase recognition problem, which is a subset of RTE. Another important similarity-based method is tree kernel (Zanzotto and Moschitti, 2006), which rely on the cross-pair similarity between two pairs (T', H') and (T'', H'').

Secondly, some approaches extract the knowledge in T-H pair and check if the knowledge in T contains the knowledge in H. Hickl (2008) transformed the T-H pair into discourse commitments, reducing the RTE task to the identification of the commitments from a T which support the inference of the H. Other works map the text to logical meaning representations, and then strict logic entailment methods, possibly by invoking theorem provers.

Thirdly, some works make use of statistical classifiers which leverages a wide variety of features. The language expression of each T-H pair are represented by a feature vector $\langle f_1, f_2 \cdots f_m \rangle$. The feature vector contains the scores of different similarity measures applied to the pair, and possibly other features.

There are also other works based on predicate-argument representations with Markov Logic for RTE, such as Rios et al. (2014) and Beltagy et al. (2013). However, they did not use discourse commitments to extract predicate-argument triples, which may lead to severe information loss.

5 Conclusion

This paper introduced a new framework to solve the Textual Entailment Recognizing task. This

framework makes full use of Markov logic network for probabilistic inference. We hold that probabilistic inference is better than strict logic method since transforming from language form to strict logic form could lose a lot of information. Therefore it is extremely hard for the theorem provers to perform well.

In addition, we use YAGO database for distant supervision. The predicates extracted from T-H pair are first aligned with YAGO. If it succeeds, the inference procedure of MLN will become much more accurate. In addition, the inference rules for constructing MLN are also extracted from YAGO database using AIME system.

This framework can correctly recognize the entailment T-H pairs which must be judged using inference. This is our improvement over the previous work.

Acknowledgement

References

Islam Beltagy, Cuong Chau, Gemma Boleda, Dan Garrette, Katrin Erk, and Raymond Mooney. 2013. Deep semantics with probabilistic logical form. In *Proceedings of the Second Joint Conference on Lexical and Computational Semantics (* SEM-13)*.

Ido Dagan, Oren Glickman, and Bernardo Magnini. 2006. The pascal recognising textual entailment challenge. In *Machine learning challenges. evaluating predictive uncertainty, visual object classification, and recognising tectual entailment*, pages 177–190. Springer.

Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, and Mark Sammons. 2006. An inference model for semantic entailment in natural language. In *Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment*, pages 261–286. Springer.

Anthony Fader, Stephen Soderland, and Oren Etzioni. 2011. Identifying relations for open information extraction. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1535–1545. Association for Computational Linguistics.

Luis Antonio Galárraga, Christina Teflioudi, Katja Hose, and Fabian Suchanek. 2013. Amie: association rule mining under incomplete evidence in ontological knowledge bases. In *Proceedings of the 22nd international conference on World Wide Web*, pages 413–422. International World Wide Web Conferences Steering Committee.

Aria D Haghighi, Andrew Y Ng, and Christopher D Manning. 2005. Robust textual inference via graph matching. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 387–394. Association for Computational Linguistics.

Andrew Hickl, John Williams, Jeremy Bensley, Kirk Roberts, Bryan Rink, and Ying Shi. 2006. Recognizing textual entailment with lccs groundhog system. In *Proceedings of the Second PASCAL Challenges Workshop*.

Andrew Hickl. 2008. Using discourse commitments to recognize textual entailment. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 337–344. Association for Computational Linguistics.

Valentin Jijkoun and Maarten de Rijke. 2005. Recognizing textual entailment using lexical similarity. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment*, pages 73–76.

Dekang Lin and Patrick Pantel. 2001. Dirt@ sbt@ discovery of inference rules from text. In *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 323–328. ACM.

Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, and Christopher D Manning. 2006. Learning to recognize features of valid textual entailments. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 41–48. Association for Computational Linguistics.

Farzaneh Mahdisoltani, Joanna Biega, and Fabian Suchanek. 2014. Yago3: A knowledge base from multilingual wikipedias. In *7th Biennial Conference on Innovative Data Systems Research*. CIDR 2015.

Prodromos Malakasiotis and Ion Androutsopoulos. 2007. Learning textual entailment using svms and string similarity measures. In *Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing*, pages 42–47. Association for Computational Linguistics.

Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on*

Natural Language Processing of the AFNLP: Volume 2-Volume 2, pages 1003–1011. Association for Computational Linguistics.

Matthew Richardson and Pedro Domingos. 2006. Markov logic networks. *Machine learning*, 62(1-2):107–136.

Miguel Rios, Lucia Specia, Alexander Gelbukh, and Ruslan Mitkov. 2014. Statistical relational learning to recognise textual entailment. In *Computational Linguistics and Intelligent Text Processing*, pages 330–339. Springer.

Marta Tatu and Dan Moldovan. 2006. A logic-based semantic approach to recognizing textual entailment. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 819–826. Association for Computational Linguistics.

Lucy Vanderwende, Arul Menezes, and Rion Snow. 2006. Microsoft research at rte-2: Syntactic contributions in the entailment task: an implementation. In *Proceedings of the Second PASCAL Challenges Workshop*.

Stephen Wan, Mark Dras, Robert Dale, and Cécile Paris. 2006. Using dependency-based features to take the para-farce out of paraphrase. In *Proceedings of the Australasian Language Technology Workshop*, volume 2006.

Andreas Wotzlaw and Ravi Coote. 2013. A logic-based approach for recognizing textual entailment supported by ontological background knowledge. *arXiv preprint arXiv:1310.4938*.

Fabio Massimo Zanzotto and Alessandro Moschitti. 2006. Automatic learning of textual entailments with cross-pair similarities. In *International conference on computational linguistics and the 44th Annual meeting of the Association for computational linguistics*, volume 1, pages 401–408. Association for Computational Linguistics.