論 文 の 内 容 の 要 旨

論文題目　　Exploiting Non-Local Information in Relation Extraction from Documents
(文書からの関係抽出における非局所的情報の利用)

氏　　名　　ラオクンラット　ナッスダ

Relation extraction from documents is one of the most common tasks in Natural Language Processing (NLP). Relations between entities are an important piece of information for deep understanding of documents. Moreover, being able to extract relations appearing in text is beneficial for various NLP applications such as textual entailment, multi-document summarization, and question answering.

Traditional machine learning-based approaches to relation extraction use only local features, i.e., features between a given pair of entities, and thus fail to incorporate useful in- formation that could be inferred from nearby entities into the classification process. In this project, in order to apply non-local information into the classification, we make use of graphs and apply a stacked learning method to classification task.

This study mainly focuses on the temporal relation classification task. However, the idea can also be applied to other NLP relation extraction tasks. In this work, the temporal relation classification task has been performed with extensive experiments to verify the proposed method.

Temporal relation classification aims to classify temporal relationships between pairs of temporal entities into one of the relation types such as BEFORE, AFTER, SIMULTANEOUS, and BEGINS. Local approaches do not consider entities that have temporal connections to the entities in the given pair at all, and thus contradictions within a document can occur. For instance, the system may predict that A happens before B, that B happens before C, and that A happens after C, which are mutually contradictory. In our model, we tackle the problem of contradictory predictions by using a stacked learning approach. The prediction for a temporal relation is made by considering the consistency of possible relations between nearby entities.

In this study, we tackle the problem of contradictory predictions by using a stacked learning approach proposed by Wolpert (1992). Stacked learning is a machine learning framework that allows one to incorporate non-local information into a structured prediction problem and has proven useful in dependency parsing (Martins et al. (2008)). We employ stacked learning in order to use the results of temporal inference as non-local features in temporal relation classification. To perform temporal inference, we use timegraphs proposed by Miller and Schubert (1990), which represent temporal connectivity of all temporal entities in each document.

Global approaches for tackling the aforementioned problem have been proposed previously (Chambers and Jurafsky (2008); Yoshikawa et al. (2009); Denis and Muller (2011); Do et al. (2012)). Chambers and Jurafsky used Integer Linear Programming (ILP) to maximize the confidence scores of the output of local classifiers in order to solve the contradictory prediction. Denis and Muller also used ILP but they enforced temporal relation coherence only on particular sets of events rather than on the entire documents. How- ever, both of the studies focused only on the temporal relations between events and used reduced sets of the temporal relations, i.e., Chambers and Jurafsky used BEFORE, AFTER, and VAGUE., while Denis and Muller used BEFORE, AFTER, OVERLAP, and NO RELATION. Do et al. employed a full set of temporal relations to construct a globally coherent timeline for an article using ILP, leveraged event coreference to support timeline construction, and associated each event with a precise time interval.

Yoshikawa et al. proposed a Markov Logic model to jointly predict the temporal relations between events and time expressions. They also used a reduced set of the relation types, i.e., BEFORE, OVERLAP, AFTER, BEFORE-OR-OVERLAP, OVERLAP-OR-AFTER, and VAGUE.

Our method differs from theirs in that their methods used transition rules to enforce consistency within each triplet of relations, but our method can also work with a set consisting of more than three relations. Moreover, in our work, the full set of temporal relations specified in TimeML are used, rather than the reduced set used in Chambers and Jurafsky (2008) and Yoshikawa et al. (2009).

Since most of the timegraph features are only applicable for multi-path TLINKs, it is important to have dense timegraphs. To alleviate the sparsity problem of timegraphs, the relation inference and the time-time connection are performed in order to increase the timegraphs' density.

We evaluate our method on the TempEval-3's Task C-relation-only data, which provides a system with all the appropriate temporal links and only needs the system to classify the relation types. The results show that by exploiting the probability values in the stacked learning approach, the classification performance improves significantly. By performing 10-fold cross validation on the Timebank corpus, we can achieve an F1 score of 60.25% based on the graph-based evaluation, which is 0.90 percentage points (*pp*) higher than that of the local approach.

We compared our system to the state-of-the-art systems that use global information in temporal relation classification and found that our system outperforms those systems. Our system can achieve 7.7 *pp* higher accuracy than Chambers's system and 0.9 *pp* higher accuracy than Yoshikawa's system. By using a stacked learning approach, we are able to include a large number of features into our models, which makes our results better than those of Yoshikawa et al. (2009), since including a large number of features into a Markov Logic model is difficult and computationally expensive.

References

- Chambers, N. and Jurafsky, D. (2008). "Jointly Combining Implicit Constraints Improves Temporal Ordering." In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP '08, pp. 698–706, Stroudsburg, PA, USA. Association for Computational Linguistics.

- Chambers, N., Wang, S., and Jurafsky, D. (2007). "Classifying temporal relations between events." In Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, ACL '07, pp. 173–176.

- Denis, P. and Muller, P. (2011). “Predicting globally-coherent temporal structures from texts via end-point inference and graph decomposition.” In IJCAI-11-International Joint Conference on Artificial Intelligence.
- Do, Q. X., Lu, W., and Roth, D. (2012). “Joint inference for event timeline construction.” In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 677–687. Association for Computational Linguistics.
- Martins, A. F. T., Das, D., Smith, N. A., and Xing, E. P. (2008). “Stacking Dependency Parsers.” In Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP ’08, pp. 157–166, Stroudsburg, PA, USA. Association for Computational Linguistics.
- Miller, S. A. and Schubert, L. K. (1990). “Time Revisited.” Comput. Intell., 6 (2), pp. 108–118.
- Wolpert, D. H. (1992). “Stacked Generalization.” Neural Networks, 5, pp. 241–259.
- Yoshikawa, K., Riedel, S., Asahara, M., and Matsumoto, Y. (2009). “Jointly Identifying Temporal Relations with Markov Logic.” 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 1 - Volume 1, ACL ’09, pp. 405–413, Stroudsburg, PA, USA. Association for Computational Linguistics.