



N-ary relations extraction From News Dataset

By

Gul Jabeen

*M.Sc in Internet Technologies and Information Systems (ITIS)
Leibniz University Hannover
Hannover, Germany*

*Project Supervisor: Dr. Besnik Fetahu
Approval Professor: Prof. Dr. Wolfgang Nejdl*

*l3s Research Center
Appelstrasse 4
30167 Hannover, Germany
Email: gul.jabeen@stud.uni-hannover.de*

May 27, 2019

Contents

1	Related work	2
2	Dataset	5
3	Inroduction	5
4	Dataset	5
5	Binary relations	5
6	N-ary relations	5

1 Related work

The prior work on relation extraction has applied on binary relations on a single sentence. Now we have worked on multiple relations (n-ary) on multiple sentences and cross sentence relation extraction. First, we will review the related work on binary relations on single sentences then later on nary relation extraction, cross sentence relation extraction, and cross sentence n-ary relation extraction with graph LSTMs.

In the binary relation extraction, the aim is to extract relations between a pair of entities within a sentence. First, [Kambhatla, 2004] has presented a statistical technique for extracting relations where they combine diverse, lexical, syntactic and semantic features. The paper [Zhou GuoDong and Min., 2005] has presented a feature-based approach for relation extraction where combine diverse, lexical, syntactic and semantic knowledge are engaged. They explored base phrase chunking information instead of using a full parse tree, which improves performance that captured all shallow information as well.

The paper [Fabian M Suchanek and Weikum, 2006] has extended the pattern matching approach for information extraction by using deep learning linguistic structures instead of shallow text patterns. Hence, they showed that how deep linguistic structures can represent for machine learning. The Paper [Yee Seng Chan, 2010] has improved the relation extraction by using background knowledge such as relationship among target relations. The paper [Oren Etzioni, 2011] has focused on identifying more meaning full relationships in a text. The paper introduced two open information extraction systems, Reverb and R2A2, which requires only shallow syntactic features and both systems double the area under precision/ recall curve compared to previous first generation open information extraction systems. In the paper [Nguyen and Grishman, 2014] has solved the problem of performance loss when the model is applied to out of domain data by evaluating word embedding and clustering an adapting feature based relation extraction systems which demonstrate the effectiveness of regularisation for the adaptability of relation extractors.

N-ary relation is a relation between more than two entities. This relation extension makes the mathematical definition of ‘relations’ a little more widely useful, and it is also a foundation for relational databases in computing. [Chinchor, 1998] has also extracted news articles using n-ary relations. The paper [Ryan McDonald and White., 2005] also applied n-ary relations to biomedical text in which they have followed two stages. First, they created a graph from pairs of related entities and then recreated the complex relations by finding maximal cliques in a graph that represents relations between the entities.

[Nicholas FitzGerald, 2015; Roth and Lapata., 2016] have also applied neural networks to semantic labelling. Following work has decomposed the nary relation into binary relation effectively by embedding the dependency path between

each pair or using feed-forward networks. This approach will help us to find more than one relation between the entities.

The cross sentence relation extraction has given advantage to many relation extraction tasks. [Swampillai and Stevenson., 2011] has used cross extraction for MUC fact and event extraction in which results demonstrates that a composite kernel approach to inter sentential relation extraction can achieve comparable results with intra sentential relation extraction. [Wick et al., 2006] also use the same technique on a different dataset; they extracted records from webpages. [Katsumasa Yoshikawa and Matsumoto., 2006] have extracted facts for biomedical domains, which incorporates co-reference relations through the concept of salience in discourse and transitivity. The paper has implemented two models, SVM pipeline and MLN joint which improves the intra sentence and cross sentence related to co-reference relations, it also leads towards h The paper [Gerber and Chai, 2010] used cross sentence relation extraction to extend the semantic role that recovers implicit inter-sentential arguments with a supervised classification model. This recovery could benefit many NLP applications.

The cross sentence relation has recently learned with distant supervision as well but that only has focused on binary relations. Many papers has been worked on relation extraction using distant supervision. This technique has applied to both binary and n-ary relation extraction.

The paper [Mike Mintz and Jurafsky., 2009] has applied distant supervision to extract binary relations which get a large number of reasonable relation with high precision patterns. The paper [Hoifung Poon and Heckerman., 2014] also extracted binary relation using distant supervision in which the technique has applied on PubMed to extract biological pathways which turned out great, they have extracted many pathways interaction in a context.

These papers [Kevin Reschke and Jurafsky., 2014; Yan Xu and Jin., 2015] have extracted n-ary relations using distant supervision, which gives better performance along with more relations. Neural networks have also been applied to distant supervision of binary relations by [Daojian Zeng and Zhao., 2015] which solves the problems of distant supervision while extracting binary relation in advance way.

The paper [Nanyun Pengl, 2017] explores the relation extraction framework using long short term memory networks (LSTM graph) in which it extracts more than two relations in multiple sentences. The framework has generalized for cross sentence n-ary relation extraction based on graph LSTM. However, the framework has evaluated into two main domain in precision medicine and contributed its effectiveness with both supervised leaning and distant learning. This approach has proved that extraction beyond sentence boundary produced far more knowledge and encoding rich linguistic knowledge provides consistent gain. This technique has been applied only to the medical field. In our thesis,

we use news dataset to apply the same techniques as in this paper.

- 2 Dataset
- 3 Inroduction
- 4 Dataset
- 5 Binary relations
- 6 N-ary relations

References

- Chinchor, N. A. (1998). Overview of muc-7/met-2. proceedings of the seventh message understanding conference (muc-7).
- Daojian Zeng, Kang Liu, Y. C. and J. Zhao. (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. in proceedings of the conference on empirical methods in natural language processing,. pp. 17–21.
- Fabian M Suchanek, G. I. and G. Weikum (2006). Combining linguistic and statistical analysis to extract relations from web documents. *In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.ACM.*, 712–717.
- Gerber, M. and J. Y. Chai (2010). Beyond nom-bank: A study of implicit arguments for nominal predicates.in proceedings of the 48th annual meeting of the association for computational linguistics, acl ’10,. *Stroudsburg, PA, USA. Association for Computational Linguistics.*, 1583–1592.
- Hoifung Poon, Chris Quirk, C. D. and D. Heckerman. (2014). Literome: Pubmed-scale genomic knowledge base in the cloud. *bioinformatics*.
- Kambhatla, N. (2004). Combining ltactic,and semantic features with maximum entropy models for extracting relations. in proceedings of the 42nd annual meeting on association for computational linguistics,demonstration sessions. *Association for Computational Linguistics*.
- Katsumasa Yoshikawa, Sebastian Riedel, T. H. M. A. and Y. Matsumoto. (2006). Coreference based event-argument relation extraction on biomedical text. *Journal of Biomedical Semantics*, 2(5):1..
- Kevin Reschke, Martin Jankowiak, M. S. C. D. M. and D. Jurafsky. (2014). Event extraction using distant supervision. in proceedings of 8th edition of the language resources and evaluation conference.
- Mike Mintz, Steven Bills, R. S. and D. 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. *Association for Computational Linguistics.*, 1003–1011.
- Nanyun Pengl, Hoifung Poon, C. Q. K. T. W.-t. Y. (2017). Cross-sentence n-ary relation extraction with graph lstms. *Transactions of the Association for Computational Linguistics (TACL) 2017* 5.
- Nguyen, T. H. and R. Grishman (2014). word representations and regularization for domain adaptation of relation extraction. *In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 68–74.
- Nicholas FitzGerald, Oscar Tackstrom, K. G. a. D. D. (2015). Semantic role labeling with neural network factors. in proceedings of the conference on empirical methods in natural language processing. in proceedings of the conference on empirical methods in natural language processing. *Lisbon, Portugal, September. Association for Computational Linguistics*, 960–970.

- Oren Etzioni, Anthony Fader, J. C. S. S. M. (2011). Open information extraction: the second generation. *Proceedings of the Twenty-Second international joint conference on Artificial Intelligence - Volume Volume One.*, 3–10.
- Roth, M. and M. Lapata. (2016). Neural semantic role labeling with dependency path embeddings. in proceedings of the 54th annual meeting of the association for computational linguistics. *Berlin, Germany, August. Association for Computational Linguistics (Volume 1: Long Papers)*, 1192–1202.
- Ryan McDonald, Fernando Pereira, S. K. S. W. Y. J. and P. White. (2005). Simple algorithms for complex relation extraction with applications to biomedical ie. *In Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics.*, 491–498.
- Swampillai, K. and M. Stevenson. (2011). Extracting relations within and across sentences. in proceedings of the conference on recent advances in natural language processing.
- Yan Xu, Lili Mou, G. L. Y. C. H. P. and Z. Jin. (2015). Classifying relations via long short term memory networks along shortest dependency paths. in proceedings of conference on empirical methods in natural language processing.
- Yee Seng Chan, D. R. (2010). Exploiting background knowledge for relation extraction. in proceedings of the 23rd international conference on computational linguistics. *Association for Computational Linguistics*, 712–717.
- Zhou GuoDong, Su Jian, Z. J. and Z. Min. (2005). Exploring various knowledge in relation extraction. in proceedings of the 43rd annual meeting on association for computational linguistics. *Association for Computational Linguistics.*, 427–434.