Improving Performance of Relation Extraction Algorithm via Leveled Adversarial PCNN and Database Expansion

Diyah Puspitaningrum
Department of Computer Science
University of Bengkulu
Bengkulu, Indonesia
diyahpuspitaningrum@gmail.com

***Abstract*—This study introduces database expansion using the Minimum Description Length (MDL) algorithm to expand the database for better relation extraction. Different from other previous relation extraction researches, our method improves system performance by expanding data. The goal of database expansion, together with a robust deep learning classifier, is to diminish wrong labels due to the incomplete or not found nature of relation instances in the relation database (e.g., Freebase). The study uses a deep learning method (Piecewise Convolutional Neural Network or PCNN) as the base classifier of our proposed approach: the leveled adversarial attention neural networks (LATTADV-ATT). In the database expansion process, the semantic entity identification is used to enlarge new instances using the most similar itemsets of the most common patterns of the data to get its pairs of entities. About the deep learning method, the use of attention of selective sentences in PCNN can reduce noisy sentences. Also, the use of adversarial perturbation training is useful to improve the robustness of system performance. The performance even further is improved using a combination of leveled strategy and database expansion. There are two issues: 1) database expansion method: rule generation by allowing step sizes on selected strong semantic of most similar itemsets with aims to find entity pair for generating instances, 2) a better classifier model for relation extraction. Experimental result has shown that the use of the database expansion is beneficial. The MDL database expansion helps improvements in all methods compared to the unexpanded method. The LATTADV-ATT performs as a good classifier with high precision P@100=0.842 (at no expansion). It is even better while implemented on the expansion data with P@100=0.891 (at expansion factor *k*=7).**

Keywords—relation extraction, database expansion, MDL, PCNN, classification

REFERENCES

[1] C. Nogueira dos Santos, B. Xiang, and B. Zhou, "Classifying Relations by Ranking with Convolutional Neural Networks," in ACL (1), 2015. [The 53rd Annual Meeting of the Association for Computational Linguistics, Beijing, China, pp.626–634, 2015].

[2] I.J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and Harnessing Adversarial Examples," CoRR abs/1412.6572: 2014.

[3] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, and D.S. Weld, "Knowledge-Based Weak Supervision for Information Extraction of Overlapping Relations," in HLT'11(1). Association for Computational Linguistics, Stroudsburg, PA, USA, pp.541–550, 2011.

[4] Y. Lin, S. Shen, Z. Liu, H. Luan, and M. Sun, "Neural Relation Extraction with Selective Attention over Instances," in Proceedings of the 54th Annual Meeting of ACL(Vol. 1: Long Papers). Association for Computational Linguistics, pp.2124–2133, 2016.

[5] T. Liu, K. Wang, B. Chang, and Z. Sui, "A Soft-label Method for Noise-tolerant Distantly Supervised Relation Extraction," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp.1790–1795, 2017.

[6] M. Mintz, S. Bills, R. Snow, and D. Jurafsky, "Distant Supervision for Relation Extraction Without Labeled Data," in Proceedings of ACL'09: Vol.2-Vol.2 (ACL'09). Association for Computational Linguistics, Stroudsburg, PA, USA, pp.1003–1011, 2009.

[7] D. Puspitaningrum, "Patterns, Models, and Queries," Ph.D. Dissertation. Department of Information and Computing Sciences, Utrecht University, 2012.

[8] D. Puspitaningrum, Fauzi, B. Susilo, J.A. Pagua, A. Erlansari, D. Andreswari, R. Efendi, and I. S. W. B. Prasetya, "An MDL-Based Frequent Itemset Hierarchical Clustering Technique to Improve Query Search Results of an Individual Search Engine," in Information Retrieval Technology, G. Zuccon, S. Geva, H. Joho, F. Scholer, A. Sun, and P. Zhang (Eds.). Springer International Publishing, Cham, pp.279–291, 2015.

[9] S. Riedel, L. Yao, and A. McCallum, "Modeling Relations and Their Mentions Without Labeled Text," in Proceedings of the 2010 ECML PKDD: Part III (ECML PKDD'10), Berlin, Heidelberg: Springer-Verlag, pp.148–163, 2010.

[10] O. Sampson and M.R. Berthold, "Widened KRIMP: Better Performance through Diverse Parallelism," in Advances in Intelligent Data Analysis XIII, Hendrik Blockeel, Matthijs van Leeuwen, and Veronica Vinciotti (Eds.), Switzerland: Springer, pp.276–285, 2014.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," in Journal of Machine Learning Research, 15, 1, pp.1929–1958, Jan 2014.

[12] M. Surdeanu, J. Tibshirani, R. Nallapati, and C.D. Manning, "Multi-instance Multi-label Learning for Relation Extraction," in Proceedings of the EMNLP-CoNLL'12. Stroudsburg, PA, USA: Association for Computational Linguistics, pp.455–465, 2012.

[13] M. van Leeuwen and D. Puspitaningrum, "Improving Tag Recommendation Using Few Associations," in Advances in Intelligent Data Analysis XI, Jaakko Hollmén, Frank Klawonn, and Allan Tucker (Eds.). Berlin, Heidelberg: Springer, pp.184–194, 2012.

[14] M. van Leeuwen, J. Vreeken, and A. Siebes, "Compression Picks Item Sets That Matter," in Knowledge Discovery in Databases: PKDD 2006, J. Fürnkranz, T. Scheffer, and M. Spiliopoulou (Eds.). Berlin, Heidelberg: Springer, pp.585–592, 2006.

[15] Y. Wu, D. Bamman, and S. Russell, "Adversarial Training for Relation Extraction," in Proceedings of the 2017 Conference on Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, pp. 1778–1783, 2017.

[16] D. Zeng, K. Liu, Y. Chen, and J. Zhao, "Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks," in EMNLP, L. Marquez, C. Callison-Burch, J. Su, D. Pighin, and Y. Marton, Eds. Association for Computational Linguistics, pp. 1753–1762, 2015.