A Rules for Weak Supervision

The pre-defined relation triple patterns adopted by RCLUS are shown in Table 4 for reference. Before applying the predefined patterns, named entity recognition will be applied for each sentence and the recognition results will be aligned with the given head and tail entity mentions for entity typing. During this step, sentences which are not recognized will not be considered for training and will be considered as negative sample if it comes from the test set.

B Experiment Setups

We conduct all the experiments on 2 NVIDIA RTX A6000 GPU's with PyTorch 1.12.1 wtih Huggingface Library [41]. For all experiments, we set the training epochs as 20 and the evaluation frequency on dev set is set to be once every 2 epochs. Best performing checkpoint on dev set is chosen for evaluation on the test set.

B.1 RCLUS Implementation Details

For pre-defined relation triple representation patterns, please refer to Appendix A.

For the clustering model's decoding function $g_l(\cdot)$ with $l \in \{h, r, t\}$, we implement them as feed-forward neural networks with each layer followed by the ReLU activation function [1]. We adopt 100, 1000, 2000, 1000, dim_l for the hidden states dimensions of each layer. The dim_l is 1024 for head/tail entity embeddings and 2048 for dependency path embeddings. For encoding mapping f, we basically reverse the layout of the three decoding functions and concatenate them to form the latent space vector $z \in \mathbb{R}^{300}$. For the clustering, the concentration parameter κ of each vMf distribution is set as 10, the λ is chosen as 5. During training, the batchsize is 256 while the learning rate is $5e-4$. Additionally, we set the tolerance threshold for optimization convergence as 0.001 which means when the ratio of the examples with changed cluster assignment is fewer than this threshold, the training will stop. The number of pre-training epochs with only objective $\mathcal{O}_{recon}$ is set as 100, the interval for updating the target assignment distribution $q(\mu_c|z_i)$ is set as 100 and the max training epochs is set as 500 which are never reached.

For prompt-tuning with sub-prompts, we used the verbalizer and the label word set from [14] except that we have modified some prompt templates and search the learning rate from $\{3e-5, 4e-5\}$ and we search the max input sequence length from $\{256, 512\}$.

The hyperparameter search space for sampling interval I is $\{1, 2, 3\}$ and the $M_{negative}$ is searched among $\{10000, 20000, 30000\}$. The currently found best combinations are $I = 3$ and $M_{negative} = 30000$

B.2 BASELINE Implementation Details

For baselines that have been originally implemented on TACRED, we adopt the same experimental setups with the original settings on TACRED.

One exception is COSINE and we search self_training_eps from $\{0.7, 0.8\}$, self_training_update_period from $\{100\}$, self_training_max_step from $\{2000, 3000\}$, and training epochs from $\{3, 20\}$. Other experimental setups are the same as the original implementation.

Relation	Head Type	Relation-Indicative Word	Tail Type
per:charges	Person	face; sentence; sentenced; sentence; sentenced; accused; charge; charged; charges; convicted; conviction; indictment; sued; sues; trial;	Criminal_charge
per:date_of_death	Person	deaths; dies; deaths; dies; dead; death; die; died; killed;	Date
per:country_of_death	Person	death; died; killed;	Country
per:cause_of_death	Person	succumbed; death; deaths; die; died; dies; killed;	Cause_of_death
org:founded_by	Organization	set; set; create; co-founder; cofounder; established; form; formed; found; founded; founder; founders; founding; founds; started;	Person
org:founded	Organization	creation; establish; form; establish; co-founder; cofounder; form; founder; co-founded; created; established; formed; founded; founding; founds; set; started;	Date
per:city_of_death	Person	murdered; murdered; death; died; dies; killed;	City
per:stateorprovince_of_death	Person	death; died; killed;	State_or_province
per:date_of_birth	Person	birth; birthday; born;	Date
per:stateorprovince_of_birth	Person	native; birth; birthday; born;	State_or_province
per:country_of_birth	Person	native; birth; birthday; birthplace; born;	Country
per:city_of_birth	Person	native; birth; birthday; birthplace; born;	City
org:shareholders	Person	acquired; bought; holding; share; holding; share; owned; stock; invest; invested; investment; investor; investors; invests; owner; owns; shareholder; shareholders; shares; stake; stakes;	Organization
per:other_family	Person	relative; sister-in-law; brother-in-law; father-in-law; mother-in-law; nephew; nephews; niece; grandfather; grandmother; grandson; granddaughter; cousin; aunts; uncle;	Person
per:title	Person	$Length$(Shortest Dependency Path) ≤ 2	Title
org:dissolved	Organization	bankruptcy; shuttering; bankruptcy; shuttering; merger; close; closed; disband; disbanded; dissolved; merged; shut; shutdown; split;takeover;	Date
per:countries_of_residence	Person	ambassador; congressman; congresswoman; lives; sen.; senator; senators; undersecretary; ambassador; ambassadors; grew; leave; return; returned; returns; apartment; citizen; emigrated; home; hometown; house; inhabitant; live; lived; living; moved; resided; resident; resides;	Country
per:stateorprovinces_of_residence	Person	sen.; senator; senators; congressman; congresswoman; grew; native; raised; represent; representative; represented; representing; sen.; senator; senators; apartment; emigrated; home; hometown; house; live; lived; lives; living; moved; resided; resident; resides;	State_or_province
per:cities_of_residence	Person	native; settled; settled; grew; move; moving; returned; apartment; emigrated; home; hometown; house; live; lived; lives; living; moved; resided; resident; resides;	City
per:religion	Person	$Length$(Shortest Dependency Path) ≤ 2	Religion
org:top_members/employees	Organization	cfo; heads; leaders; leads; vice-chairman; cfo; governor; heads; leaders; leadership; vice; vice-chairman; vice-governor; ceo; chairman; chairwoman; chief; commander; coo; director; head; lead; leader; led; president; vice-president;	Person
org:number_of_employees/members	Organization	employs; employs; workforce; employees; members; membership; people; workers;	Number
per:schools_attended	Person	attending; education; studied; studies; education; educated; graduation; students; study; studying; attended; attends; degree; degrees; doctorate; graduate; graduated; graduates; graduating; schooled; schools; student	Organization
per:employee_of	Person	boss; chief; directors; driver; executive; heads; leader; leaders; manager; officer; officials; professor; undersecretary; worker; works; chief; attorney; co-chairman; correspondent; driver; elected; executive; fellow; minister;officer; officials; professor; serve; speaker; undersecretary; appointed; chairman; chairwoman; director; employed; employee; employees; head; hired; member; members; official; president; secretary; served; spokesman; spokeswoman; worked; workers;	Organization
per:siblings	Person	sibling; siblings; brother; brothers; sister; sisters;	Person
per:spouse	Person	marrying; divorce; marrying; wedding; ex-wife; husband; marriage; married; spouse; wife;	Person
org:stateorprovince_of_headquarters	Organization	base; based; located; founded; firm; firms; office; headquarters; headquartered; belongs;	State_or_province
org:country_of_headquarters	Organization	agency; headquarters; company; companies; based; root; firm; founded; located;	Country
org:city_of_headquarters	Organization	base; located; based; headquarters; headquartered; company; firm; office; founded;	City
org:stateorprovince_of_branch	Organization	branches; run; runs; base; based; in; operate; operated; operates; office; site; headquarters; headquartered; plant; office; company; firm; firms; branch; operations; operators; operator;	State_or_province
org:country_of_branch	Organization	branch; branches; in; office; offices; base; based; headquarters; headquartered; run; runs; operations; operate; operated; operates; operators; operator; agency; company; head; firm; subsidiary; companies; plants; site; sites;	Country
org:city_of_branch	Organization	branch; branches; site; run; runs; in; operations; operates; operate; operated; base; based; headquarters; headquartered; plant; office; offices; company; firm; firms; subsidiary;	City

Table 4. Pre-defined relation patterns of RCLUS. The patterns are designed based on example sentences and the relation definitions from *TAC KBP 2014 Slot Descriptions* [12]. Note that the joint relation-indicative words are in black. Additional relation-indicative words for TACRED are in red. Additional relation-indicative words for TACREV are in blue. Additional relation-indicative words for ReTACRED are in cyan.

References

1. Agarap, A.F.: Deep learning using rectified linear units (relu). CoRR **abs/1803.08375** (2018), http://arxiv.org/abs/1803.08375
2. Aina, L., Gulordava, K., Boleda, G.: Putting words in context: LSTM language models and lexical ambiguity. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 3342–3348. Association for Computational Linguistics, Florence, Italy (Jul 2019). https://doi.org/10.18653/v1/P19-1324, https://aclanthology.org/P19-1324
3. Alt, C., Gabryszak, A., Hennig, L.: TACRED revisited: A thorough evaluation of the TACRED relation extraction task. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. pp. 1558–1569. Association for Computational Linguistics, Online (Jul 2020). https://doi.org/10.18653/v1/2020.acl-main.142, https://aclanthology.org/2020.acl-main.142
4. Banerjee, A., Dhillon, I.S., Ghosh, J., Sra, S.: Clustering on the unit hypersphere using von mises-fisher distributions. J. Mach. Learn. Res. **6**, 1345–1382 (2005)
5. Batista, D.S., Martins, B., Silva, M.J.: Semi-supervised bootstrapping of relationship extractors with distributional semantics. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 499–504. Association for Computational Linguistics, Lisbon, Portugal (Sep 2015). https://doi.org/10.18653/v1/D15-1056, https://aclanthology.org/D15-1056
6. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 1877–1901. Curran Associates, Inc. (2020), https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
7. Chen, Y.N., Hakkani-Tür, D., Tur, G.: Deriving local relational surface forms from dependency-based entity embeddings for unsupervised spoken language understanding. In: 2014 IEEE Spoken Language Technology Workshop (SLT). pp. 242–247 (2014). https://doi.org/10.1109/SLT.2014.7078581
8. Curran, J., Murphy, T., Scholz, B.: Minimising semantic drift with mutual exclusion bootstrapping. Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics pp. 172–180 (12 2008)
9. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the em algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**(1), 1–38 (1977), http://www.jstor.org/stable/2984875
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://aclanthology.org/N19-1423
11. Ding, N., Chen, Y., Han, X., Xu, G., Xie, P., Zheng, H., Liu, Z., Li, J.Z., Kim, H.G.: Prompt-learning for fine-grained entity typing. ArXiv **abs/2108.10604** (2021)
12. Ellis, J., Getman, J., Strassel, S.M.: Overview of linguistic resources for the tac kbp 2014 evaluations: Planning, execution, and results. In: Proceedings of TAC KBP 2014 Workshop, National Institute of Standards and Technology. pp. 17–18 (2014)

13. Fundel, K., Küffner, R., Zimmer, R.: Relex - relation extraction using dependency parse trees. Bioinformatics **23 3**, 365–71 (2007)
14. Han, X., Zhao, W., Ding, N., Liu, Z., Sun, M.: Ptr: Prompt tuning with rules for text classification. AI Open **3**, 182–192 (2022). https://doi.org/https://doi.org/10.1016/j.aiopen.2022.11.003, https://www.sciencedirect.com/science/article/pii/S2666651022000183
15. Hancock, B., Varma, P., Wang, S., Bringmann, M., Liang, P., Ré, C.: Training classifiers with natural language explanations. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 1884–1895. Association for Computational Linguistics, Melbourne, Australia (Jul 2018). https://doi.org/10.18653/v1/P18-1175, https://aclanthology.org/P18-1175
16. Hinton, G.E., Zemel, R.S.: Autoencoders, minimum description length and helmholtz free energy. In: Proceedings of the 6th International Conference on Neural Information Processing Systems. p. 3–10. NIPS'93, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1993)
17. Hu, S., Ding, N., Wang, H., Liu, Z., Li, J.Z., Sun, M.: Knowledgeable prompt-tuning: Incorporating knowledge into prompt verbalizer for text classification. In: Annual Meeting of the Association for Computational Linguistics (2021)
18. Lu, Y., Bartolo, M., Moore, A., Riedel, S., Stenetorp, P.: Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. CoRR **abs/2104.08786** (2021), https://arxiv.org/abs/2104.08786
19. van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of Machine Learning Research **9**(86), 2579–2605 (2008), http://jmlr.org/papers/v9/vandermaaten08a.html
20. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., McClosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations. pp. 55–60. Association for Computational Linguistics, Baltimore, Maryland (Jun 2014). https://doi.org/10.3115/v1/P14-5010, https://aclanthology.org/P14-5010
21. Mausam, Schmitz, M., Soderland, S., Bart, R., Etzioni, O.: Open language learning for information extraction. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 523–534. Association for Computational Linguistics, Jeju Island, Korea (Jul 2012), https://aclanthology.org/D12-1048
22. Meng, Y., Huang, J., Wang, G., Zhang, C., Zhuang, H., Kaplan, L., Han, J.: Spherical text embedding. In: Advances in Neural Information Processing Systems (2019)
23. Meng, Y., Shen, J., Zhang, C., Han, J.: Weakly-supervised neural text classification. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. p. 983–992. CIKM '18, Association for Computing Machinery, New York, NY, USA (2018). https://doi.org/10.1145/3269206.3271737, https://doi.org/10.1145/3269206.3271737
24. Meng, Y., Zhang, Y., Huang, J., Zhang, Y., Zhang, C., Han, J.: Hierarchical topic mining via joint spherical tree and text embedding. In: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. p. 1908–1917. KDD '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3394486.3403242, https://doi.org/10.1145/3394486.3403242
25. Nakashole, N., Weikum, G., Suchanek, F.: PATTY: A taxonomy of relational patterns with semantic types. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. pp. 1135–1145. Association for Computational Linguistics, Jeju Island, Korea (Jul 2012), https://aclanthology.org/D12-1104

26. Nayak, T., Majumder, N., Goyal, P., Poria, S.: Deep neural approaches to relation triplets extraction: a comprehensive survey. Cognitive Computation **13**, 1215 – 1232 (2021)
27. Qu, M., Ren, X., Zhang, Y., Han, J.: Weakly-supervised relation extraction by pattern-enhanced embedding learning. In: Proceedings of the 2018 World Wide Web Conference. p. 1257–1266. WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE (2018). https://doi.org/10.1145/3178876.3186024, https://doi.org/10.1145/3178876.3186024
28. Ratner, A., Bach, S.H., Ehrenberg, H., Fries, J., Wu, S., Ré , C.: Snorkel: Rapid training data creation with weak supervision. Proceedings of the VLDB Endowment **11**(3), 269–282 (nov 2017). https://doi.org/10.14778/3157794.3157797, https://doi.org/10.14778%2F3157794.3157797
29. Ratner, A., Sa, C.D., Wu, S., Selsam, D., Ré, C.: Data programming: Creating large training sets, quickly. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. p. 3574–3582. NIPS'16, Curran Associates Inc., Red Hook, NY, USA (2016)
30. Ren, W., Li, Y., Su, H., Kartchner, D., Mitchell, C., Zhang, C.: Denoising multi-source weak supervision for neural text classification. In: Findings of the Association for Computational Linguistics: EMNLP 2020. pp. 3739–3754. Association for Computational Linguistics, Online (Nov 2020). https://doi.org/10.18653/v1/2020.findings-emnlp.334, https://aclanthology.org/2020.findings-emnlp.334
31. Schick, T., Schmid, H., Schütze, H.: Automatically identifying words that can serve as labels for few-shot text classification. In: Proceedings of the 28th International Conference on Computational Linguistics. pp. 5569–5578. International Committee on Computational Linguistics, Barcelona, Spain (Online) (Dec 2020). https://doi.org/10.18653/v1/2020.coling-main.488, https://aclanthology.org/2020.coling-main.488
32. Shin, T., Razeghi, Y., IV, R.L.L., Wallace, E., Singh, S.: Eliciting knowledge from language models using automatically generated prompts. ArXiv **abs/2010.15980** (2020)
33. Shwartz, V., Goldberg, Y., Dagan, I.: Improving hypernymy detection with an integrated path-based and distributional method. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 2389–2398. Association for Computational Linguistics, Berlin, Germany (Aug 2016). https://doi.org/10.18653/v1/P16-1226, https://aclanthology.org/P16-1226
34. Simmons, R.F.: Answering english questions by computer: A survey. Commun. ACM **8**(1), 53–70 (jan 1965). https://doi.org/10.1145/363707.363732, https://doi.org/10.1145/363707.363732
35. Socher, R., Karpathy, A., Le, Q.V., Manning, C.D., Ng, A.Y.: Grounded compositional semantics for finding and describing images with sentences. Transactions of the Association for Computational Linguistics **2**, 207–218 (2014). https://doi.org/10.1162/tacl_a_00177, https://aclanthology.org/Q14-1017
36. Stoica, G., Platanios, E.A., P'oczos, B.: Re-tacred: Addressing shortcomings of the tacred dataset. In: AAAI Conference on Artificial Intelligence (2021)
37. Varma, P., Ré, C.: Snuba: Automating weak supervision to label training data. Proc. VLDB Endow. **12**(3), 223–236 (nov 2018). https://doi.org/10.14778/3291264.3291268, https://doi.org/10.14778/3291264.3291268
38. Wang, C., Kalyanpur, A., Fan, J., Boguraev, B.K., Gondek, D.C.: Relation extraction and scoring in deepqa. IBM Journal of Research and Development **56**(3.4), 9:1–9:12 (2012). https://doi.org/10.1147/JRD.2012.2187239
39. Wang, H., Liu, B., Li, C., Yang, Y., Li, T.: Learning with noisy labels for sentence-level sentiment classification. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural

Language Processing (EMNLP-IJCNLP). pp. 6286–6292. Association for Computational Linguistics, Hong Kong, China (Nov 2019). https://doi.org/10.18653/v1/D19-1655, https://aclanthology.org/D19-1655

40. Wang, H., Tian, F., Gao, B., Zhu, C., Bian, J., Liu, T.Y.: Solving verbal questions in iq test by knowledge-powered word embedding. In: Conference on Empirical Methods in Natural Language Processing (2015)
41. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A.: Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. pp. 38–45. Association for Computational Linguistics, Online (Oct 2020). https://doi.org/10.18653/v1/2020.emnlp-demos.6, https://aclanthology.org/2020.emnlp-demos.6
42. Xie, J., Girshick, R., Farhadi, A.: Unsupervised deep embedding for clustering analysis. In: Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. p. 478–487. ICML'16, JMLR.org (2016)
43. Xu, Y., Mou, L., Li, G., Chen, Y., Peng, H., Jin, Z.: Classifying relations via long short term memory networks along shortest dependency paths. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. pp. 1785–1794. Association for Computational Linguistics, Lisbon, Portugal (Sep 2015). https://doi.org/10.18653/v1/D15-1206, https://aclanthology.org/D15-1206
44. Xue, F., Sun, A., Zhang, H., Chng, E.S.: Gdpnet: Refining latent multi-view graph for relation extraction. In: AAAI Conference on Artificial Intelligence (2020)
45. Yu, Y., Zuo, S., Jiang, H., Ren, W., Zhao, T., Zhang, C.: Fine-tuning pre-trained language model with weak supervision: A contrastive-regularized self-training approach. ArXiv **abs/2010.07835** (2020)
46. Zhang, J., Yu, Y., Li, Y., Wang, Y., Yang, Y., Yang, M., Ratner, A.: WRENCH: A comprehensive benchmark for weak supervision. In: Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2) (2021), https://openreview.net/forum?id=Q9SKS5k8io
47. Zhang, Y., Zhong, V., Chen, D., Angeli, G., Manning, C.D.: Position-aware attention and supervised data improve slot filling. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. pp. 35–45. Association for Computational Linguistics, Copenhagen, Denmark (Sep 2017). https://doi.org/10.18653/v1/D17-1004, https://aclanthology.org/D17-1004
48. Zhou, W., Lin, H., Lin, B.Y., Wang, Z., Du, J., Neves, L., Ren, X.: Nero: A neural rule grounding framework for label-efficient relation extraction. In: Proceedings of The Web Conference 2020. p. 2166–2176. WWW '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3366423.3380282, https://doi.org/10.1145/3366423.3380282
49. Zhuang, L., Wayne, L., Ya, S., Jun, Z.: A robustly optimized BERT pre-training approach with post-training. In: Proceedings of the 20th Chinese National Conference on Computational Linguistics. pp. 1218–1227. Chinese Information Processing Society of China, Huhhot, China (Aug 2021), https://aclanthology.org/2021.ccl-1.108