just entities. Last, we provide a distant supervision framework for generating the training data so as to make our approach scalable.

7 CONCLUSION

In this paper, we introduced the knowledge graph fact contextualization task and proposed NFCM, a weakly-supervised method to address it. NFCM first generates a candidate set for a query fact by looking at 1 or 2-hop neighbors and then ranks the candidate facts using supervised machine learning. NFCM combines handcrafted features with features that are automatically identified using deep learning. We use distant supervision to boost the gathering of training data by using a large entity-tagged text corpus that has a high overlap with entities in the KG we use. Our experimental results show that (i) distant supervision is an effective means for gathering training data for this task, (ii) NFCM significantly outperforms several heuristic baselines for this task, and (iii) both the handcrafted and automatically-learned features contribute to the retrieval effectiveness of NFCM. For future work, we aim to explore more sophisticated ways of combining handcrafted with automatically learned features for ranking. Additionally, we want to explore other data sources for gathering training data, such as news articles and click logs. Finally, we want to explore methods for combining and presenting the ranked facts in search engine result pages in a diversified fashion.

Data

To facilitate reproducibility of our results, we share the data used to run our experiments at https://www.techatbloomberg.com/research-weakly-supervised-contextualization-knowledge-graph-facts/.

Acknowledgements

The authors would like to thank the anonymous reviewers (and especially reviewer #1) for their useful and constructive feedback. This research was supported by Ahold Delhaize, Amsterdam Data Science, the Bloomberg Research Grant program, the China Scholarship Council, the Criteo Faculty Research Award program, Elsevier, the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 312827 (VOX-Pol), the Google Faculty Research Awards program, the Microsoft Research Ph.D. program, the Netherlands Institute for Sound and Vision, the Netherlands Organisation for Scientific Research (NWO) under project nrs CI-14-25, 652.002.001, 612.001.551, 652.001.003, and Yandex. All content represents the opinion of the authors, which is not necessarily shared or endorsed by their respective employers and/or sponsors.

REFERENCES

[1] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. TensorFlow: A System for Large-Scale Machine Learning.. In *OSDI*, Vol. 16. USENIX Association, 265–283.

[2] B. Aleman-Meza, C. Halaschek-Weiner, I. B. Arpinar, Cartic Ramakrishnan, and A. P. Sheth. 2005. Ranking complex relationships on the semantic Web. *IEEE Internet Computing* 9, 3 (May 2005), 37–44.

[3] Enrique Alfonseca, Katja Filippova, Jean-Yves Delort, and Guillermo Garrido. 2012. Pattern Learning for Relation Extraction with a Hierarchical Topic Model. In *ACL*. ACL, Stroudsburg, PA, USA, 54–59.

[4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural Machine Translation by Jointly Learning to Align and Translate. *CoRR* abs/1409.0473 (2014), 1–15.

[5] Hannah Bast, Buchhold Björn, and Elmar Haussmann. 2016. Semantic search on text and knowledge bases. *Found. Trends Inf. Retr.* 10, 2-3 (June 2016), 119–271.

[6] Roi Blanco, Berkant Barla Cambazoglu, Peter Mika, and Nicolas Torzec. 2013. Entity Recommendations in Web Search. In *ISWC*. Springer Berlin Heidelberg, Berlin, Heidelberg, 33–48.

[7] Roi Blanco, Giuseppe Ottaviano, and Edgar Meij. 2015. Fast and Space-Efficient Entity Linking for Queries. In *WSDM*. ACM, New York, NY, USA, 179–188.

[8] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge. In *SIGMOD*. ACM, New York, NY, USA, 1247–1250.

[9] Alexey Borisov, Pavel Serdyukov, and Maarten de Rijke. 2016. Using metafeatures to increase the effectiveness of latent semantic models in web search. In *WWW*. ACM, New York, NY, USA, 1081–1091.

[10] Horatiu Bota, Ke Zhou, and Joemon M. Jose. 2016. Playing Your Cards Right: The Effect of Entity Cards on Search Behaviour and Workload. In *CHIIR*. ACM, New York, NY, USA, 131–140.

[11] M. Cho, K. Alahari, and J. Ponce. 2013. Learning Graphs to Match. In *ICCV*. IEEE, 25–32.

[12] Rajarshi Das, Arvind Neelakantan, David Belanger, and Andrew McCallum. 2017. Chains of Reasoning over Entities, Relations, and Text using Recurrent Neural Networks. In *EACL*. ACL, Valencia, Spain, 132–141.

[13] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W Bruce Croft. 2017. Neural ranking models with weak supervision. In *SIGIR*. ACM, New York, NY, USA, 65–74.

[14] Lujun Fang, Anish Das Sarma, Cong Yu, and Philip Bohannon. 2011. Rex: explaining relationships between entity pairs. *VLDB* 5, 3 (2011), 241–252.

[15] Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. 2017. Creating Training Corpora for NLG Micro-Planners. In *ACL*. ACL, 179–188.

[16] Kelvin Guu, John Miller, and Percy Liang. 2015. Traversing Knowledge Graphs in Vector Space. In *EMNLP*. ACL, 318–327.

[17] Faegheh Hasibi, Krisztian Balog, and Svein Erik Bratsberg. 2017. Dynamic Factual Summaries for Entity Cards. In *SIGIR*. ACM, New York, NY, USA, 773–782.

[18] Jizhou Huang, Wei Zhang, Shiqi Zhao, Shiqiang Ding, and Haifeng Wang. 2017. Learning to Explain Entity Relationships by Pairwise Ranking with Convolutional Neural Networks. In *IJCAI*. IJCAI, 4018–4025.

[19] Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *CoRR* abs/1412.6980 (2014), 1–15. arXiv:1412.6980

[20] Rémi Lebret, David Grangier, and Michael Auli. 2016. Neural Text Generation from Structured Data with Application to the Biography Domain. In *EMNLP*. ACL, 1203–1213.

[21] Xiao Ling and Daniel S. Weld. 2012. Fine-grained Entity Recognition. In *AAAI*. AAAI Press, 94–100.

[22] Iris Miliaraki, Roi Blanco, and Mounia Lalmas. 2015. From "Selena Gomez" to "Marlon Brando": Understanding Explorative Entity Search. In *WWW*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 765–775.

[23] Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *ACL/AFNLP*. ACL, 1003–1011.

[24] Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. 2016. A review of relational machine learning for knowledge graphs: From multi-relational link prediction to automated knowledge graph construction. *Proc. of the IEEE* 104, 1 (2016), 11–33.

[25] Thomas Pellissier Tanon, Denny Vrandečić, Sebastian Schaffert, Thomas Steiner, and Lydia Pintscher. 2016. From Freebase to Wikidata: The Great Migration. In *WWW*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland, 1419–1428.

[26] Giuseppe Pirrò. 2015. Explaining and Suggesting Relatedness in Knowledge Graphs. In *ISWC*. Springer-Verlag New York, Inc., New York, NY, USA, 622–639.

[27] Xiang Ren, Ahmed El-Kishky, Chi Wang, Fangbo Tao, Clare R. Voss, and Jiawei Han. 2015. ClusType: Effective Entity Recognition and Typing by Relation Phrase-Based Clustering. In *KDD*. ACM, New York, NY, USA, 995–1004.

[28] Sebastian Riedel, Limin Yao, and Andrew McCallum. 2010. Modeling Relations and Their Mentions Without Labeled Text. In *ECML-PKDD*. Springer-Verlag, Berlin, Heidelberg, 148–163.

[29] Stephan Seufert, Klaus Berberich, Srikanta J. Bedathur, Sarath Kumar Kondreddi, Patrick Ernst, and Gerhard Weikum. 2016. ESPRESSO: Explaining Relationships Between Entity Sets. In *CIKM*. ACM, New York, NY, USA, 1311–1320.

[30] Daniil Sorokin and Iryna Gurevych. 2017. Context-Aware Representations for Knowledge Base Relation Extraction. In *EMNLP*. ACL, 1784–1789.

[31] Mihai Surdeanu, Julie Tibshirani, Ramesh Nallapati, and Christopher D. Manning. 2012. Multi-instance Multi-label Learning for Relation Extraction. In *EMNLP-CoNLL*. ACL, 455–465.

[32] Nikos Voskarides, Edgar Meij, and Maarten de Rijke. 2017. Generating Descriptions of Entity Relationships. In *ECIR*. Springer International Publishing, Cham, 317–330.

[33] Nikos Voskarides, Edgar Meij, Manos Tsagkias, Maarten de Rijke, and Wouter Weerkamp. 2015. Learning to Explain Entity Relationships in Knowledge Graphs. In *ACL-IJCNLP*. ACL, 564–574.

[34] Wen-tau Yih, Ming-Wei Chang, Xiaodong He, and Jianfeng Gao. 2015. Semantic Parsing via Staged Query Graph Generation: Question Answering with Knowledge Base. In *ACL-IJCNLP*. ACL, 1321–1331.