been used at all in this work, is likely to play a factor in this prediction task. For our own task(s), extending to more features, such as additional QPPs, can further enhance the prediction performance. Further ahead, predicting the type and semantics of the reformulation can help satisfy consumers' needs more rapidly and effectively, as reformulation plays such a central role in e-commerce search.

Acknowledgements This paper is based upon work supported in part by the Israel Science Foundation under grant no. 433/12, the German Research Foundation (DFG) via the German-Israeli Project Cooperation (DIP, grant DA 1600/1-1), and an eBay grant.

REFERENCES

[1] Martin Arlitt. 2000. Characterizing web user sessions. *SIGMETRICS Perform. Eval. Rev.* 28, 2 (2000), 50–63.
[2] Ahmed Hassan Awadallah, Ranjitha Gurunath Kulkarni, Umut Ozertem, and Rosie Jones. 2015. Characterizing and predicting voice query reformulation. In *Proc. of CIKM.* 543–552.
[3] Ahmed Hassan Awadallah, Xiaolin Shi, Nick Craswell, and Bill Ramsey. 2013. Beyond clicks: query reformulation as a predictor of search satisfaction. In *Proc. of CIKM.* 2019–2028.
[4] Ahmed Hassan Awadallah, Ryen W. White, Susan T. Dumais, and Yi-Min Wang. 2014. Struggling or exploring?: disambiguating long search sessions. In *Proc. of WSDM.* 53–62.
[5] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. 2017. *Deep learning.*
[6] Paolo Boldi, Francesco Bonchi, Carlos Castillo, Debora Donato, Aristides Gionis, and Sebastiano Vigna. 2008. The query-flow graph: model and applications. In *Proc. of CIKM.* 609–618.
[7] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
[8] Eliot Brenner, Jun Zhao, Aliasgar Kutiyanawala, and Zheng Yan. 2018. End-to-end neural ranking for eCommerce product search: An application of task models and textual embeddings. In *The SIGIR 2018 Workshop On eCommerce.*
[9] David Carmel and Elad Yom-Tov. 2010. *Estimating the query difficulty for information retrieval.* Morgan & Claypool Publishers.
[10] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM TIST* 2, 3 (2011), 27.
[11] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. In *Proc. of KDD.* 785–794.
[12] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine Learning* 20, 3 (1995), 273–297.
[13] Steve Cronen-Townsend, Yun Zhou, and W. Bruce Croft. 2002. Predicting query performance. In *Proc. of SIGIR.* 299–306.
[14] Ronan Cummins, Joemon M. Jose, and Colm O'Riordan. 2011. Improved query performance prediction using standard deviation. In *Proc. of SIGIR.* 1089–1090.
[15] Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. 2018. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proc. of ACL.* 1383–1392.
[16] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5 (1971), 378–382.
[17] Jerome H Friedman. 2001. Greedy function approximation: A gradient boosting machine. *Annals of statistics* (2001), 1189–1232.
[18] Katharina C Furtner, Thomas Mandl, and Christa Womser-Hacker. 2015. Effects of Auto-Suggest on the Usability of Search in eCommerce.. In *ISI.* 178–190.
[19] Sreenivas Gollapudi, Samuel Ieong, and Anitha Kannan. 2012. Structured query reformulations in commerce search. In *Proc. of CIKM.* 1890–1894.
[20] Ido Guy and Kira Radinsky. 2017. Structuring the unstructured: From startup to making sense of eBay's huge eCommerce inventory. In *Proc. of SIGIR.* 1351.
[21] Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. 2018. Mix 'n match: Integrating text matching and product substitutability within product search. In *Proc. of CIKM.* 1373–1382.
[22] Claudia Hauff, Leif Azzopardi, and Djoerd Hiemstra. 2009. The combination and evaluation of query performance prediction methods. In *Proc. of ECIR.* 301–312.
[23] Yujing Hu, Qing Da, Anxiang Zeng, Yang Yu, and Yinghui Xu. 2018. Reinforcement learning to rank in e-Commerce search engine: Formalization, analysis, and application. In *Proc. of SIGKDD.* 368–377.
[24] Jeff Huang and Efthimis N. Efthimiadis. 2009. Analyzing and evaluating query reformulation strategies in web search logs. In *Proc. of CIKM.* 77–86.
[25] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. 2009. Patterns of query reformulation during Web searching. *JASIST* 60, 7 (2009), 1358–1371.
[26] Jiepu Jiang, Ahmed Hassan Awadallah, Xiaolin Shi, and Ryen W. White. 2015. Understanding and predicting graded search satisfaction. In *Proc. of WSDM.* 57–66.
[27] Jyun-Yu Jiang, Yen-Yu Ke, Pao-Yu Chien, and Pu-Jen Cheng. 2014. Learning user reformulation behavior for query auto-completion. In *Proc. of SIGIR.* 445–454.
[28] Rosie Jones and Kristina Lisa Klinkner. 2008. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *Proc. of CIKM.* 699–708.
[29] Li-Jen Kao and Yo-Ping Huang. 2017. Predicting purchase intention according to fan page users' sentiment. In *Proc. of SMC.* 831–835.
[30] Elad Kravi, Ido Guy, Avihai Mejer, David Carmel, Yoelle Maarek, Dan Pelleg, and Gilad Tsur. 2016. One Query, Many Clicks: Analysis of Queries with Multiple Clicks by the Same User. In *Proc. of CIKM.* 1423–1432.
[31] Rohan Kumar, Mohit Kumar, Neil Shah, and Christos Faloutsos. 2018. Did we get it right? predicting query performance in e-Commerce search. In *The SIGIR 2018 Workshop On eCommerce.*
[32] Or Levi, Ido Guy, Fiana Raiber, and Oren Kurland. 2018. Selective Cluster Presentation on the Search Results Page. *ACM TOIS* 36, 3, Article 28 (2018), 42 pages.
[33] Saurav Manchanda, Mohit Sharma, and George Karypis. 2019. Intent term selection and refinement in e-commerce queries. *CoRR* abs/1908.08564 (2019).
[34] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Proc. of NIPS.* 3111–3119.
[35] Eric W Noreen. 1989. *Computer-intensive methods for testing hypotheses.* Wiley New York.
[36] Daan Odijk, Ryen W. White, Ahmed Hassan Awadallah, and Susan T. Dumais. 2015. Struggling and success in web search. In *Proc. of CIKM.* 1551–1560.
[37] Jay M. Ponte and W. Bruce Croft. 1998. A language modeling approach to information retrieval. In *Proc. of SIGIR.* 275–281.
[38] Fiana Raiber and Oren Kurland. 2014. Query-performance prediction: setting the expectations straight. In *Proc. of SIGIR.* 13–22.
[39] Hadas Raviv, Oren Kurland, and David Carmel. 2014. Query performance prediction for entity retrieval. In *Proc. of SIGIR.* 1099–1102.
[40] Stephen E. Robertson, Steve Walker, Susan Jones, Micheline Hancock-Beaulieu, and Mike Gatford. 1994. Okapi at TREC-3. In *Proc. of TREC-3.*
[41] Haggai Roitman, Shai Erera, Oren Sar Shalom, and Bar Weiner. 2017. Enhanced mean retrieval score estimation for query performance prediction. In *Proc. of ICTIR.* 35–42.
[42] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. 2017. On application of learning to rank for e-Commerce search. In *Proc. of SIGIR.* 475–484.
[43] Nikos Sarkas, Stelios Paparizos, and Panayiotis Tsaparas. 2010. Structured annotations of web queries. In *Proc. of SIGMOD.* 771–782.
[44] Humphrey Sheil, Omer Rana, and Ronan Reilly. 2018. Predicting purchasing intent: Automatic feature learning using recurrent neural networks. In *The SIGIR 2018 Workshop On eCommerce.*
[45] Anna Shtok, Oren Kurland, and David Carmel. 2016. Query performance prediction using reference lists. *ACM Trans. Inf. Syst.* 34, 4 (2016), 19:1–19:34.
[46] Anna Shtok, Oren Kurland, David Carmel, Fiana Raiber, and Gad Markovits. 2012. Predicting query performance by query-drift estimation. *ACM Transactions on Information Systems* 30, 2 (2012), 11.
[47] Gyanit Singh, Nish Parikh, and Neel Sundaresn. 2011. User behavior in zero-recall ecommerce queries. In *Proc. of SIGIR.* 75–84.
[48] Parikshit Sondhi, Mohit Sharma, Pranam Kolari, and ChengXiang Zhai. 2018. A taxonomy of queries for e-Commerce search. In *Proc. of SIGIR.* 1245—-1248.
[49] Ning Su, Jiyin He, Yiqun Liu, Min Zhang, and Shaoping Ma. 2018. User Intent, behaviour, and perceived satisfaction in product search. In *Proc. of WSDM.* 547–555.
[50] Zehong Tan, Canran Xu, Mengjie Jiang, Hua Yang, and Xiaoyuan Wu. 2017. Query rewrite for null and low search results in eCommerce. In *The SIGIR 2017 Workshop On eCommerce.*
[51] Yongquan Tao and Shengli Wu. 2014. Query performance prediction by considering score magnitude and variance together. In *Proc. of CIKM.* 1891–1894.
[52] Arthur Toth, Louis Tan, Giuseppe Di Fabbrizio, and Ankur Datta. 2017. Predicting shopping behavior with mixture of RNNs. In *The SIGIR 2017 Workshop On eCommerce.*
[53] Hen Tzaban, Ido Guy, Asnat Greenstein-Messica, Arnon Dagan, Lior Rokach, and Bracha Shapira. 2020. Product Bundle Identification using Semi-Supervised Learning. In *Proc. of SIGIR.*
[54] Teng Xiao, Jiaxin Ren, Zaiqiao Meng, Huan Sun, and Shangsong Liang. 2019. Dynamic bayesian metric learning for personalized product search. In *Proc. of CIKM.* 1693–1702.
[55] Yingwei Xin, Ethan Hart, Vibhuti Mahajan, and Jean-David Ruvini. 2018. Learning better internal structure of words for sequence labeling. In *Proc. of EMNLP.* 2584–2593.
[56] Yuan Zhang, Dong Wang, and Yan Zhang. 2019. Neural IR meets graph embedding: A ranking model for product search. In *Proc. of WWW.* 2390–2400.
[57] Ying Zhao, Falk Scholer, and Yohannes Tsegay. 2008. Effective pre-retrieval query performance prediction using similarity and variability evidence. In *Proc. of ECIR.* 52–64.
[58] Yun Zhou and W. Bruce Croft. 2007. Query performance prediction in web search environments. In *Proc. of SIGIR.* 543–550.