

framework over the traditional hinge loss training approach. This confirms the result in [25].

Combining features vs. combining scores: One of our initial intuitions is that it is better to combine representation and interaction features than combining the scores they produce. This can be confirmed by comparing HM-MLP and HM-sum (equivalent to the Duet model) in Table 4: It is clear that combining features is better than combining scores. This suggests that MLP can indeed make better use of the features when they are presented together, and this allows them to have possible interactions.

Flexibility of ILM: Our ILM is also general, it is possible to replace the representation/interaction module with different ones. For interaction module, we replace the pyramid-based one with the histogram-based one depicted in Fig. 2 and build ILM-Hist. Experimental results in Table 4 show that it could still outperform traditional baselines and offer comparable performance with respect to the original ILM.

Learning curve: Finally, we plot the NDCG@10 curve of ILM-Letor+ on the validation set of fold 1 of MQ2007 in Fig. 4, together with BM25, LM for comparison. We can observe that at the beginning of the training, the NDCG value first increases. As training goes on, the performance reaches the maximum and begins to decrease. This could be possibly due to overfitting or to the lack of sufficient training data. The curve shows an area in the middle where our proposed model can outperform traditional models.

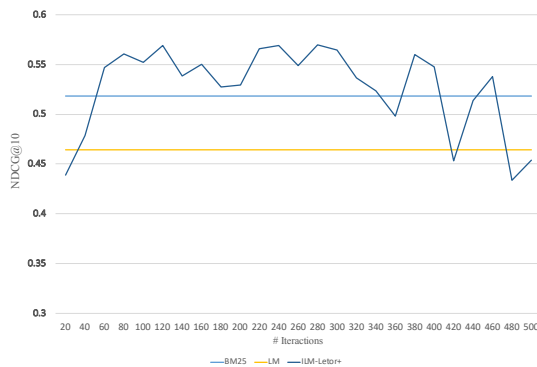


Figure 4: Learning curve of nDCG@10 on Validation data on MQ2007

5 CONCLUSION AND FUTURE WORK

In this paper, we proposed an integrated learning framework to integrate both representation- and interaction-based features. In addition, we also integrate feature learning with ranking-function learning. Experiments on public datasets confirm the effectiveness of integrated learning of ranking features of different nature and ranking function.

In this study, the interactions between different matching mechanisms happen on the feature level. In future studies, we plan to explore their interactions on intermediate layers and test the ILM model on large-scale datasets such as MSMARCO.

REFERENCES

- [1] Qingyao Ai, Xuanhui Wang, Nadav Golbandi, Michael Bendersky, and Marc Najork. 2018. Learning Groupwise Scoring Functions Using Deep Neural Networks. *CoRR* abs/1811.04415 (2018). arXiv:1811.04415
- [2] Christopher JC Burges. 2010. From ranknet to lambdarank to lambdamart: An overview. *Learning* 11, 23-581 (2010), 81.
- [3] Yixing Fan, Jiafeng Guo, Yanyan Lan, Jun Xu, Chengxiang Zhai, and Xueqi Cheng. 2018. Modeling Diverse Relevance Patterns in Ad-hoc Retrieval. In *SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*. 375–384.
- [4] Jian Gu, Guang-Hua Yin, Pengfei Huang, Jinlu Guo, and Lijun Chen. 2017. An improved back propagation neural network prediction model for subsurface drip irrigation system. *Computers & Electrical Engineering* 60 (2017), 58–65.
- [5] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W. Bruce Croft. 2016. A Deep Relevance Matching Model for Ad-hoc Retrieval. In *CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. 55–64.
- [6] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional Neural Network Architectures for Matching Natural Language Sentences. In *NIPS 2014, December 8-13 2014, Montreal, Quebec, Canada*. 2042–2050.
- [7] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry P. Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013*. 2333–2338.
- [8] Robert Krovetz. 2000. Viewing morphology as an inference process. *Artif. Intell.* 118, 1-2 (2000), 277–294.
- [9] Tie-Yan Liu. 2009. Learning to Rank for Information Retrieval. *Foundations and Trends in Information Retrieval* 3, 3 (2009), 225–331. <https://doi.org/10.1561/1500000016>
- [10] Tie-Yan Liu. 2011. *Learning to Rank for Information Retrieval*. Springer. <https://doi.org/10.1007/978-3-642-14267-3>
- [11] Bhaskar Mitra and Nick Craswell. 2019. An Updated Duet Model for Passage Re-ranking. *CoRR* abs/1903.07666 (2019).
- [12] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. 2017. Learning to Match using Local and Distributed Representations of Text for Web Search. In *WWW 2017, Perth, Australia, April 3-7, 2017*. 1291–1299.
- [13] Yifan Nie, Alessandro Sordani, and Jian-Yun Nie. 2018. Multi-level Abstraction Convolutional Model with Weak Supervision for Information Retrieval. In *SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018*. 985–988.
- [14] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, and Xueqi Cheng. 2016. A Study of MatchPyramid Models on Ad-hoc Retrieval. *CoRR* abs/1606.04648 (2016).
- [15] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2016. Text Matching as Image Recognition. In *AAAI 2016, February 12-17, 2016, Phoenix, Arizona, USA*. 2793–2799.
- [16] Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Jingfang Xu, and Xueqi Cheng. 2017. DeepRank: A New Deep Architecture for Relevance Ranking in Information Retrieval. In *CIKM 2017, Singapore, November 06 - 10, 2017*. 257–266.
- [17] Rama Kumar Pasumarthi, Xuanhui Wang, Cheng Li, Sebastian Bruch, Michael Bendersky, Marc Najork, Jan Pfeifer, Nadav Golbandi, Rohan Anil, and Stephan Wolf. 2018. TF-Ranking: Scalable TensorFlow Library for Learning-to-Rank. *CoRR* abs/1812.00073 (2018). arXiv:1812.00073
- [18] Stephen E. Robertson. 2000. Evaluation in Information Retrieval. In *ESSIR 2000, Varenna, Italy, September 11-15, 2000, Revised Lectures*. 81–92.
- [19] Stephen E. Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends in Information Retrieval* 3, 4 (2009), 333–389.
- [20] Aliaksei Severyn and Alessandro Moschitti. 2015. Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. In *SIGIR 2015, Santiago, Chile, August 9-13, 2015*. 373–382.
- [21] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning semantic representations using convolutional neural networks for web search. In *WWW'14, Seoul, Republic of Korea, April 7-11, 2014*. 373–374.
- [22] Richard Socher, Danqi Chen, Christopher D. Manning, and Andrew Y. Ng. 2013. Reasoning With Neural Tensor Networks for Knowledge Base Completion. In *NIPS 2013, Lake Tahoe, Nevada, United States*. 926–934.
- [23] Ming Tan, Tian Xia, Lily Guo, and Shaojun Wang. 2013. Direct optimization of ranking measures for learning to rank models. In *SIGKDD 2013, Chicago, IL, USA, August 11-14, 2013*. 856–864.
- [24] Shengxian Wan, Yanyan Lan, Jun Xu, Jiafeng Guo, Liang Pang, and Xueqi Cheng. 2016. Match-SRNN: Modeling the Recursive Matching Structure with Spatial RNN. In *IJCAI 2016, New York, NY, USA, 9-15 July 2016*. 2922–2928.
- [25] Xuanhui Wang, Cheng Li, Nadav Golbandi, Michael Bendersky, and Marc Najork. 2018. The LambdaLoss Framework for Ranking Metric Optimization. In *CIKM 2018, Torino, Italy, October 22-26, 2018*. 1313–1322.
- [26] Yining Wang, Liwei Wang, Yuanzhi Li, Di He, and Tie-Yan Liu. 2013. A Theoretical Analysis of NDCG Type Ranking Measures. In *COLT 2013, Princeton University, NJ, USA*. 25–54.