

Table 5: nDCG@10 of Representative Queries for Inter Models

Topic_num	Query	Inter-S0	Inter-S2	Inter-MultiMatch
54	President of the United States	0.1370	0.0367	0.1429
153	Pocono	0.0609	0.0595	0.0658
145	vines for shade	0.2225	0.3333	0.4201
155	last supper painting	0.2240	0.3479	0.3938

model with only term-level S_0 matching outperforms the model with high-level score S_2 . The latter model may expand too much the semantic of the query, which performs poorly. For example among the documents retrieved for the query “President of the United States” by *Inter – S2*, a number of documents are about presidents of other countries than the United States which are irrelevant for this query. This is an example of over generalization. For conceptual queries such as “vines for shade”, “last supper painting”, the model with high-level score S_2 outperforms the model with only term-level score S_0 , since it captures the high-level interactions required for these queries. For example, the documents retrieved for “last supper painting” by *Inter – S2* include the desired documents about “description of the last supper painting”, “significance of last supper painting in Catholicism”. In this case, the query has been correctly generalized at conceptual level. In both above cases, by dynamically combining the matching scores of the 3 levels of abstraction by a gating mechanism, our Multi-Match-Inter model can outperform each of the single-level base models. These examples show the ability of our model to use the appropriate level(s) of matching depending on the query.

6 CONCLUSION AND FUTURE WORK

Representation-based and interaction-based IR models have been proposed in several previous studies, but they have been trained and evaluated on different datasets, which makes it difficult to compare their performance on a fair basis. In this paper, we build convolutional models on both schemas, train them and evaluate their performance on the same training and testing collection. Our first goal is to compare the two approaches as fairly as possible. The experimental results clearly show that interaction-based models always outperform representation-based models. The promising performance of interaction-based model is partly attributed to their capacity to learn the local interaction patterns rather than learning global representations in representation-based model. Our observation is not new, and it is consistent with the previous studies. However, our contribution is to compare the models on a fair basis. This allows us to draw a clear conclusion that interaction-based models are more appropriate than representation-based models.

Moreover, to achieve query-dependent matching, we integrate multi-level matching mechanism into both representation- and interaction-based models and evaluate their performance. The experimental results show that in both schemas, multi-level matching models could outperform models with single-level matching. This result shows the usefulness to design models that can cope with low-level lexical matching as well as high-level semantic matching. In addition, we observe that with multi-level matching mechanism,

representation-based model still performs worse than interaction-based model, which further confirms that it is more preferable to employ interaction-based models for ad hoc search.

In this study, we have imposed several constraints in our implementation and experiments in order to compare models as fairly as possible. We deliberately left several useful options such as combining the two schemas, combining with a traditional ranking model (e.g. BM25) in a learning-to-rank framework, and so on. These constraints are also the limitations of this study: we have not tested all the possible options in our comparison. In our future work, we will extend the comparison to models that use more options.

REFERENCES

- [1] Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W. Bruce Croft. 2017. Neural Ranking Models with Weak Supervision. In *Proceedings of ACM SIGIR '17, Shinjuku, Tokyo, Japan, August 7-11, 2017*. 65–74.
- [2] Alex Graves, Abdel-rahman Mohamed, and Geoffrey E. Hinton. 2013. Speech recognition with deep recurrent neural networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*. 6645–6649.
- [3] Jiafeng Guo, Yixing Fan, Qingyao Ai, and W. Bruce Croft. 2016. A Deep Relevance Matching Model for Ad-hoc Retrieval. In *Proceedings of ACM CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. 55–64.
- [4] Baotian Hu, Zhengdong Lu, Hang Li, and Qingcai Chen. 2014. Convolutional Neural Network Architectures for Matching Natural Language Sentences. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*. 2042–2050.
- [5] 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 *Proceedings of CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013*. 2333–2338.
- [6] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* 20, 4 (2002), 422–446.
- [7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*. 1106–1114.
- [8] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. 2017. Learning to Match using Local and Distributed Representations of Text for Web Search. In *Proceedings of WWW 2017, Perth, Australia, April 3-7, 2017*. 1291–1299.
- [9] Gia-Hung Nguyen, Laure Soulier, Lynda Tamine, and Nathalie Bricon-Souf. 2017. DSRIM: A Deep Neural Information Retrieval Model Enhanced by a Knowledge Resource Driven Representation of Documents. In *Proceedings of ACM ICTIR 2017, Amsterdam, The Netherlands, October 1-4, 2017*. 19–26.
- [10] 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).
- [11] Aliaksei Severyn and Alessandro Moschitti. 2015. Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks. In *Proceedings of ACM SIGIR '15 Santiago, Chile, August 9-13, 2015*. 373–382.
- [12] Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. Learning semantic representations using convolutional neural networks for web search. In *Proceedings of WWW '14, Seoul, Republic of Korea, April 7-11, 2014, Companion Volume*. 373–374.
- [13] Chenyan Xiong, Jamie Callan, and Tie-Yan Liu. 2017. Word-Entity Duet Representations for Document Ranking. In *Proceedings of ACM SIGIR '17, Shinjuku, Tokyo, Japan, August 7-11, 2017*. 763–772.
- [14] Liu Yang, Qingyao Ai, Jiafeng Guo, and W. Bruce Croft. 2016. aNMM: Ranking Short Answer Texts with Attention-Based Neural Matching Model. In *Proceedings of ACM CIKM 2016, Indianapolis, IN, USA, October 24-28, 2016*. 287–296.
- [15] Wenpeng Yin and Hinrich Schütze. 2015. MultiGranCNN: An Architecture for General Matching of Text Chunks on Multiple Levels of Granularity. In *Proceedings of ACL 2015, July 26-31, 2015, Beijing, China, Volume 1: Long Papers*. 63–73.
- [16] Yisong Yue, Thomas Finley, Filip Radlinski, and Thorsten Joachims. 2007. A support vector method for optimizing average precision. In *Proceedings of ACM SIGIR '07, Amsterdam, The Netherlands, July 23-27, 2007*. 271–278.
- [17] Xiaoqiang Zhou, Baotian Hu, Jiaxin Lin, Yang Xiang, and Xiaolong Wang. 2015. ICR-HIT: A Deep Learning based Comment Sequence Labeling System for Answer Selection Challenge. In *Proceedings of the 9th International Workshop on Semantic Evaluation, Denver, Colorado, USA, June 4-5, 2015*. 210–214.