Inf Retrieval J
DOI 10.1007/s10791-017-9323-9

Neural information retrieval: introduction to the special issue

Nick Craswell[1] · W. Bruce Croft[2] · Maarten de Rijke[3] · Jiafeng Guo[4] · Bhaskar Mitra[1,5]

Received: 30 October 2017 / Accepted: 6 November 2017

1 Introduction

The applications of neural network models, shallow or deep, to information retrieval (IR) tasks falls under the purview of *neural IR*. Over the years, machine learning methods—including neural networks—have been popularly employed in IR, such as in learning-to-rank (LTR) frameworks (Liu 2009). Recently, neural representation learning and neural models with deep architectures have demonstrated significant improvements in speech recognition, machine translation, and computer vision tasks (LeCun et al. 2015). Similar methods are now being explored by the IR community that may lead to new models and performance breakthroughs for retrieval scenarios.

In text retrieval, neural approaches may refer to the application of pre-trained term embeddings for ad-hoc retrieval (Ai et al. 2016; Diaz et al. 2016; Ganguly et al. 2015; Guo et al. 2016b; Kenter et al. 2016; Mitra et al. 2016; Roy et al. 2016; Vulić and Moens 2015; Zamani and Croft 2016a, b), or training deep neural networks end-to-end for the ranking task (Cohen and Croft 2016; Guo et al. 2016a; Huang et al. 2013; Mitra et al. 2017; Nanni et al. 2017; Pang et al. 2016a, b; Severyn and Moschitti 2015). But neural IR may also encompass the use of neural networks in proactive recommendations (Luukkonen et al. 2016; Van Gysel et al. 2017), query formulation and suggestions (Mitra 2015; Mitra and Craswell 2015; Sordoni et al. 2015), modelling user behaviour (Borisov et al. 2016a, b),

✉ Bhaskar Mitra
bmitra@microsoft.com

1 Microsoft, Redmond, WA, USA

2 University of Massachusetts, Boston, MA, USA

3 University of Amsterdam, Amsterdam, Netherlands

4 Chinese Academy of Sciences, Beijing, China

5 University College London, London, UK

Published online: 10 November 2017

entity ranking (Van Gysel et al. 2016a, b), conversational agents (Yan et al. 2016; Zhou et al. 2016), and multi-modal retrieval (Ma et al. 2015). The growing body of work in this area has been supplemented by an increasing number of recent workshops (Craswell et al. 2016a, b, 2017) and tutorials (Kenter et al. 2017; Li and Lu 2016; Mitra and Craswell 2017a, b). This special issue of the Information Retrieval journal provides an additional venue for the findings from research happening at the intersection of information retrieval and neural networks.

2 Overview of papers

We received 11 submissions in response to the call-for-papers first announced in August 2016. We present the four papers that were accepted from this pool. The first paper serves as a survey of the current body of literature in neural IR. The other three papers are focused on the application of neural models for retrieval in the context of text, image, and music, respectively. We briefly discuss these papers in the remainder of this section.

Onal et al. (2017) summarize the large body of current work in the area of neural IR. The survey starts with an overview of basic neural network concepts and popular models for text processing tasks. The body of this work then focuses on providing a broad taxonomy of neural models for different IR tasks and scenarios. The survey concludes with a reflection on lessons learned and potential future directions for the field.

Yang et al. (2017) investigate the use of pre-trained term embeddings for a Twitter classification task. They find that better performance is achieved when the data for training the text embeddings aligns with the classification dataset. They evaluate various hyper-parameter choices, including context window sizes and size of the embedding vectors, and report their findings and insights.

Carrara et al. (2017) focus on image retrieval based on short text queries. They propose multiple deep architectures that jointly learn representations of images and text queries for ranking, and report state-of-the art performance.

Wang et al. (2017) explore neural models for context-aware music recommendation. They propose the *music2vec* model for learning distributed representation of songs, and use the learnt representation in the recommendation task. The proposed model significantly outperforms baseline methods, particularly when the data is sparse.

3 Conclusion

The papers included in this special issue cover a broad range of IR scenarios spanning over text and multimedia retrieval. In addition, the survey paper provides a useful summary of current work that we believe will serve as useful reference for research in this area. Despite the growing body of work in neural IR, it is important to keep in mind that we are still in the early days of this field. The community is ready for exciting new breakthroughs, but we must also ground our enthusiasm in thorough empirical evaluation of these new methods and push for deeper understanding of their relationships to classical IR approaches in future work.

Acknowledgements We thank Charles Clarke, co-editor of the journal, for his guidance during the editing of this special issue. We further like to thank all the authors who submitted their work and the reviewers who spent time and effort providing invaluable feedback in the making of this special issue.

References

Ai, Q., Yang, L., Guo, J., & Croft, W.B. (2016). Analysis of the paragraph vector model for information retrieval. In *Proceedings of the ICTIR* (pp. 133–142). ACM.

Borisov, A., Markov, I., de Rijke, M., & Serdyukov, P. (2016a). A context-aware time model for web search. In *Proceedings of the 39th international ACM SIGIR conference on research and development in information retrieval* (pp. 205–214). ACM.

Borisov, A., Markov, I., de Rijke, M., & Serdyukov, P. (2016b). A neural click model for web search. In *Proceedings of the WWW* (pp. 531–541).

Carrara, F., Esuli, A., Fagni, T., Falchi, F., & Fernández, A.M. (2017). Picture it in your mind: Generating high level visual representations from textual descriptions. *Information Retrieval Journal*. https://doi.org/10.1007/s10791-017-9318-6.

Cohen, D., & Croft, W.B. (2016). End to end long short term memory networks for non-factoid question answering. In *Proceedings of the ICTIR* (pp. 143–146). ACM.

Craswell, N., Croft, W.B., Guo, J., Mitra, B., & de Rijke, M. (2016a). Neu-ir: The sigir 2016 workshop on neural information retrieval. In *Proceedings of the SIGIR*, ACM.

Craswell, N., Croft, W. B., Guo, J., Mitra, B., & de Rijke, M. (2016b). Report on the sigir 2016 workshop on neural information retrieval (neu-ir). *ACM Sigir forum*, *50*(2), 96–103.

Craswell, N., Croft, W.B., de Rijke, M., Guo, J., & Mitra, B. (2017). Neu-ir'17: Neural information retrieval. In *Proceedings of the SIGIR*, ACM.

Diaz, F., Mitra, B., & Craswell, N. (2016). Query expansion with locally-trained word embeddings. In *Proceedings of the ACL.*

Ganguly, D., Roy, D., Mitra, M., & Jones, G.J. (2015). Word embedding based generalized language model for information retrieval. In *Proceedings of the SIGIR* (pp. 795–798). ACM.

Guo, J., Fan, Y., Ai, Q., & Croft, W.B. (2016a). A deep relevance matching model for ad-hoc retrieval. In *Proceedings of the CIKM* (pp. 55–64). ACM.

Guo, J., Fan, Y., Ai, Q., & Croft, W.B. (2016b). Semantic matching by non-linear word transportation for information retrieval. In *Proceedings of the CIKM* (pp. 701–710). ACM.

Huang, P.S., He, X., Gao, J., Deng, L., Acero, A., & Heck, L. (2013). Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the CIKM* (pp. 2333–2338). ACM.

Kenter, T., Borisov, A., & de Rijke, M. (2016). *Siamese cbow: Optimizing word embeddings for sentence representations*. arXiv:160604640

Kenter, T., Borisov, A., Van Gysel, C., Dehghani, M., de Rijke, M., & Mitra, B. (2017). Neural networks for information retrieval (nn4ir). In *Proceedings of the SIGIR*, ACM.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

Li, H., & Lu, Z. (2016). Deep learning for information retrieval. In *Proceedings of the SIGIR*, ACM.

Liu, T. Y. (2009). Learning to rank for information retrieval. *Foundation and Trends in Information Retrieval*, *3*(3), 225–331.

Luukkonen, P., Koskela, M., & Floréen, P. (2016). *Lstm-based predictions for proactive information retrieval*. arXiv:160606137

Ma, L., Lu, Z., Shang, L., & Li, H. (2015). Multimodal convolutional neural networks for matching image and sentence. In *Proceedings of the IEEE international conference on computer vision* (pp. 2623–2631).

Mitra, B. (2015). Exploring session context using distributed representations of queries and reformulations. In *Proceedings of the SIGIR* (pp. 3–12). ACM.

Mitra, B., & Craswell, N. (2015). Query auto-completion for rare prefixes. In *Proceedings of the CIKM*, ACM.

Mitra, B., & Craswell, N. (2017a). An introduction to neural information retrieval. Foundations and Trends® in Information Retrieval (to appear).

Mitra, B., & Craswell, N. (2017b). Neural text embeddings for information retrieval. In *Proceedings of the WSDM* (pp. 813–814). ACM.

Mitra, B., Nalisnick, E., Craswell, N., & Caruana, R. (2016). *A dual embedding space model for document ranking*. arXiv:160201137

Mitra, B., Diaz, F., & Craswell, N. (2017). Learning to match using local and distributed representations of text for web search. In *Proceedings of the WWW* (pp. 1291–1299).

Nanni, F., Mitra, B., Magnusson, M., & Dietz, L. (2017). Benchmark for complex answer retrieval. In *Proceedings of the ICTIR*, ACM.

Onal, K.D., Zhang, Y., Altingovde, I.S., Rahman, M.M., et al. (2017). Neural information retrieval: At the end of the early years. *Information Retrieval Journal*. https://doi.org/10.1007/s10791-017-9321-y.

Pang, L., Lan, Y., Guo, J., Xu, J., & Cheng, X. (2016a). *A study of matchpyramid models on ad-hoc retrieval*. arXiv:160604648

Pang, L., Lan, Y., Guo, J., Xu, J., Wan, S., & Cheng, X. (2016b). Text matching as image recognition. In *Proceedings of the AAAI*.

Roy, D., Paul, D., Mitra, M., & Garain, U. (2016). *Using word embeddings for automatic query expansion*. arXiv:160607608.

Severyn, A., & Moschitti, A. (2015). Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the SIGIR* (pp. 373–382). ACM.

Sordoni, A., Bengio, Y., Vahabi, H., Lioma, C., Simonsen, J.G., & Nie, J.Y. (2015). *A hierarchical recurrent encoder-decoder for generative context-aware query suggestion*. arXiv:150702221.

Van Gysel, C., de Rijke, M., & Kanoulas, E. (2016a). Learning latent vector spaces for product search. In*Proceedings of the CIKM* (pp. 165–174). ACM.

Van Gysel, C., de Rijke, M., & Worring, M. (2016b). Unsupervised, efficient and semantic expertise retrieval. In *Proceedings of the 25th international conference on world wide web, international world wide web conferences steering committee* (pp. 1069–1079).

Van Gysel, C., Mitra, B., Venanzi, M., Rosemarin, R., Kukla, G., Grudzien, P., & Cancedda, N. (2017). Reply with: Proactive recommendation of email attachments. In *Proceedings of the CIKM*.

Vulić, I., & Moens, M.F. (2015). Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In *Proceedings of the SIGIR* (pp. 363–372). ACM.

Wang, D., Deng, S., & Xu, G. (2017). Sequence-based context-aware music recommendation. *Information Retrieval Journal*. https://doi.org/10.1007/s10791-017-9317-7.

Yan, R., Song, Y., & Wu, H. (2016). Learning to respond with deep neural networks for retrieval-based human-computer conversation system. In *Proceedings of the SIGIR* (pp. 55–64). ACM.

Yang, X., Macdonald, C., & Ounis, I. (2017). Using word embeddings in twitter election classification. *Information Retrieval Journal*. https://doi.org/10.1007/s10791-017-9319-5.

Zamani, H., & Croft, W.B. (2016a). Embedding-based query language models. In *Proceedings of the ICTIR* (pp. 147–156). ACM.

Zamani, H., & Croft, W.B. (2016b). Estimating embedding vectors for queries. In *Proceedings of the ICTIR* (pp. 123–132). ACM.

Zhou, X., Dong, D., Wu, H., Zhao, S., Yan, R., Yu, D., Liu, X., & Tian, H. (2016). Multi-view response selection for human-computer conversation. In *Proceedings of EMNLP16* (pp. 372–381).