Report on the SIGIR 2016 Workshop on Neural Information Retrieval (Neu-IR)

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

[1]Microsoft [2]University of Massachusetts [3]Chinese Academy of Sciences [4]University of Amsterdam

Abstract

The SIGIR 2016 workshop on Neural Information Retrieval (Neu-IR) took place on 21 July, 2016 in Pisa. The goal of the Neu-IR (pronounced "*New IR*") workshop was to serve as a forum for academic and industrial researchers, working at the intersection of information retrieval (IR) and machine learning, to present new work and early results, compare notes on neural network toolkits, share best practices, and discuss the main challenges facing this line of research. In total, 19 papers were presented, including oral and poster presentations. The workshop program also included a session on invited "lightning talks" to encourage participants to share personal insights and negative results with the community. The workshop was well-attended with more than 120 registrations.

1 Introduction

In recent years, deep neural networks (DNNs) have yielded significant performance improvements on speech recognition and computer vision tasks [12, 18], as well as led to exciting breakthroughs in novel application areas such as automatic voice translation [19], image captioning [10, 38], and conversational agents [37]. Despite demonstrating good performance on natural language processing (NLP) tasks (e.g., language modelling [14] and machine translation [2]), the performance of DNNs on information retrieval (IR) tasks has had relatively less scrutiny. Recent work in this area has mainly focused on applications of word embeddings [11, 25, 42] and learning embeddings for short text [13, 16, 17, 24, 33, 35, 36], while there has also some work on using deep neural networks to model user interaction behavior [4, 5, 23].

The relative lack of positive results in this area of information retrieval is partially due to the fact that IR tasks, such as document ranking, are fundamentally different from NLP tasks, but also because the IR and neural network communities are only beginning to focus on the application of these techniques to core information retrieval problems. Given that deep learning has had such a big impact, first on speech processing and computer vision and now, increasingly, also on computational linguistics, it seems possible that deep learning will have a major impact on information retrieval and therefore this was an ideal time for a workshop in this area.

The first international **Neu-IR** (pronounced "*new IR*") workshop was a forum for new research relating to deep learning and other neural network based approaches to IR. The purpose was to provide an opportunity for people to present new work and early results, compare notes on neural network toolkits, share best practices, and discuss the main challenges facing this line of research. The workshop received a significantly large response, both in terms of submissions (close to 30) and registrations (more than 120).

2 Scope and format

The Neu-IR workshop was a gathering of academic and industrial researchers working at the intersection of IR and neural networks. We solicited [8] submission of papers of two to six pages, representing reports of original research, preliminary research results, proposals for new work, descriptions of neural network based toolkits tailored for IR, and position papers. Papers presented at the workshop were required to be uploaded to `https://arXiv.org` but were considered non-archival, and may be submitted elsewhere (modified or not). The workshop site maintains a link to the arXiv versions. This submission policy was adopted to make the workshop a forum for the presentation and discussion of current work, without preventing the work from being published elsewhere. We solicited submissions relevant to the following main themes:

- The application of neural network models in IR tasks, including but not limited to:
  - Full text document retrieval, passage retrieval, question answering
  - Web search, searching social media, distributed information retrieval, entity ranking
  - Learning to rank combined with neural network based representation learning
  - User / task modelling, personalization, diversity
  - Query formulation assistance, query recommendation, conversational search
  - Multimedia retrieval
- Fundamental modelling challenges faced in such applications, including but not limited to:
  - Learning dense representations for long documents
  - Dealing with rare queries and rare words
  - Modelling text at different granularities (character, word, passage, document)
  - Compositionality of vector representations
  - Jointly modelling queries, documents, entities and other structured/knowledge data
- Best practices for research and development in the area, dealing with concerns such as:
  - Finding sufficient publicly-available training data
  - Baselines, test data, avoiding overfitting
  - Neural network toolkits[1]
  - Real-world use cases, deployment at scale

The workshop featured a mix of different presentations formats, including oral presentations, poster presentations, "lightning talks", and invited keynotes. The full program and schedule of the workshop is available on the workshop website.[2]

[1] `http://deeplearning.net/software_links/`

[2] `https://www.microsoft.com/en-us/research/event/neuir2016`

3 Keynotes

The workshop featured two invited keynotes. The opening keynote was given by Tomas Mikolov – currently a research scientist at Facebook AI Research and previously a member of the Google Brain team, where he developed and implemented efficient algorithms for computing distributed representations of words (word2vec project) – which has seen several recent applications in IR [11, 26, 42]. The second keynote was given by Hang Li, director of the Noahs Ark Lab of Huawei Technologies, adjunct professors of Peking University and Nanjing University, and an ACM Distinguished Scientist.

3.1 Recurrent Networks and Beyond

In this talk,[3] Mikolov presented recurrent neural networks, first introducing the basic architecture and then presenting a brief history of key research developments. He then focused on extensions relating to the network's ability to model longer-term context. For example, this can be used to bring term repetition patterns more in line with observed repetition in a language modeling task. He then introduced the problem of learning algorithmic patterns, which is not well solved by standard deep learning techniques. This led to a discussion of long term research plans, and the possibility of developing true artificial intelligence, which requires the definition of appropriate goals and data sets.

3.2 Does IR Need Deep Learning?

In recent years, deep learning has become the key technology of state-of-the-art systems in areas of computer science, such as computer vision, speech processing, and natural language processing. A question that naturally arises is whether deep learning will also become important in information retrieval. In fact, there has been a large amount of effort made to address the question and significant progress has been achieved. Yet there is still doubt about whether it is the case.

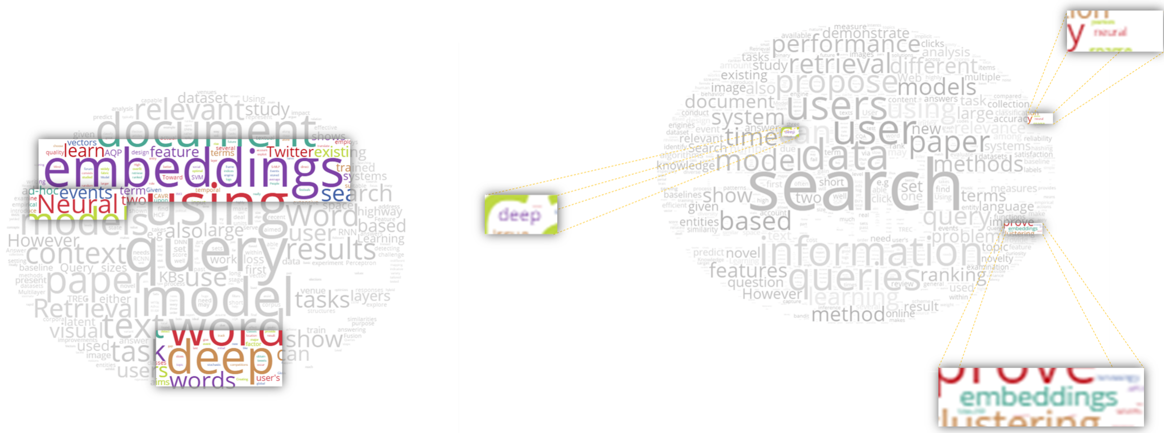
In this talk,[4] Li argued that, if we take a broad view on IR, then deep learning can indeed greatly boost IR. It has already been observed that deep learning can make great improvements on some hard problems in IR such as question answering from knowledge bases, image retrieval, etc. On the other hand, for some traditional IR tasks, in some sense easy tasks, such as document retrieval, the improvements might not be so notable. Li introduced some of the work on deep learning for IR conducted at Huawei Noahs Ark Lab, to support his claim. He also discussed the strengths and limitations of deep learning, IR problems on which deep learning can potentially make significant contributions, as well as future directions of research on IR.

4 Accepted papers

The workshop received 27 submissions, excluding three incomplete submissions. Every paper was reviewed by at least two members of the program committee and finally 19 submissions were accepted (acceptance rate of 70%). Among the accepted papers, there were a few

[3] http://www.slideshare.net/BhaskarMitra3/recurrent-networks-and-beyond-by-tomas-mikolov

[4] http://www.hangli-hl.com/uploads/3/4/4/6/34465961/does_ir_need_deep_learning.pdf



(a) Neu-IR 2016 word cloud

(b) SIGIR 2016 word cloud

Figure 1: Comparing key themes across the papers presented at the Neu-IR 2016 workshop and the ones presented in the main SIGIR 2016 conference. The word cloud summaries presented here were generated using `http://www.wordle.net`.

popular themes. 8 papers [1, 3, 9, 22, 30–32, 41] were related to word embeddings. 10 papers [6, 7, 15, 20, 21, 27–29, 34, 39] focused on the application of deep neural networks to different IR tasks. The accepted papers also covered a broad range of tasks, including question/answering [34], proactive IR [21], knowledge-based IR [27], conversational models [29], text-to-image [6], and document ranking [1, 7, 9, 28, 31, 32, 40]. Figure 1 visually contrasts some of the common themes that emerged among the workshop papers against the core topics covered by the full papers presented in the main SIGIR 2016 conference.

Geographically, the accepted papers (based on first author) ranged from 9 countries and 3 continents – four each from France and India, two each from China, Denmark, UK, and USA, and finally one each from Austria, Finland, and Italy. Based on the first authors affiliation, two of the accepted papers came from the industry and the rest from academia.

All 19 papers were presented as posters. Additionally, five papers [9, 22, 28–30] were selected for oral presentations considering reviewer feedback, relevance to the main focus of the workshop, and diversity of topics.

5 Lessons from the trenches

The Lessons from the Trenches session was aimed at encouraging researchers who are actively working in the intersection of IR and neural networks to share their insights with the broader community. In particular, we were interested to hear about,

- Key challenges faced in making neural models work effectively for IR tasks
- Best practices and related insights
- Negative results

The session comprised of seven "lightning talks" – each speaker presented a single slide

Table 1: Speakers and topics for the "Lessons from the Trenches" session.

Speaker	Topic
Qingyao Ai	Learning representations for ad-hoc retrieval
Debasis Ganguly	Inverted list organization of quantized word and doc vectors
Alessandro Moschitti	Combining kernels and neural networks
Jun Xu	Relevance $\neq$ similarity: lessons from search result diversification
Grady Simon	Passage relevance with dual-input architectures
Bhaskar Mitra	Think sparse, act dense
Sergey I. Nikolenko	Topic quality metrics based on word embeddings

within the allotted three minutes, with an additional two minutes reserved for questions. Table 1 lists all the speakers, and the respective topics covered, during this session.

6 Group discussion

The final session of the day was a group discussion among all attendees. The goal of the discussion was to identify key challenges and opportunities in the area of neural IR.

A popular topic during this session focused on the lack of positive results from deep neural network (DNN) models on the ad-hoc document retrieval tasks. One view from participants was that there is insufficient training data, and larger data will be required before DNN models can succeed on document ranking. However, given the lack of literature in this area, it was not immediately obvious what the appropriate dataset should contain, and how to start making reasonable progress towards building this dataset. Some attendees posited that progress in the area of deep learning may push forward less explored retrieval tasks, such as conversational and proactive IR. There seemed to be a clear consensus that it was important for the IR community to start focusing on designing appropriate evaluation methods and metrics for these recently emerging retrieval tasks.

Finally, there seemed to be an agreement that deep neural network models are likely to have a strong influence on the area of IR in the next few years. However, in spite of the excitement in the IR community about this area, it was not directly obvious how, if at all, the core SIGIR conference should evolve to better attract papers on neural IR in next year's conference. A workshop like Neu-IR, in the meantime, could play a vital role in bridging the gap between the IR and the neural network communities.

7 Conclusions

During the opening keynote of the SIGIR 2016 conference, Christopher Manning predicted a significant influx of deep neural network related papers for IR in the next few years. However, he encouraged the community to be mindful of some of the "irrational exuberance" that plagues the field today. The first SIGIR workshop on neural information retrieval received an unexpectedly high number of submissions and registrations. These are clear indications that the IR community is excited by the recent developments in the area of deep neural networks.

However, there is still a lack of clear understanding about how important these new machine learning approaches will be when addressing traditional and emerging IR tasks, and how they can (or should) co-exist with traditional IR approaches. This is indeed an exciting time for this area of research and we believe that besides attempting to simply demonstrate empirical progress on retrieval tasks, our explorations with neural models should also provide new insights about IR itself. In return, we should also look for opportunities to apply IR intuitions into improving these neural models, and their application to non-IR tasks.

8 Acknowledgements

The workshop organizers are sincerely grateful to all the program committee members – Carsten Eickhoff (ETH Zurich), Debasis Ganguly (Dublin City University), Katja Hoffman (Microsoft Research), Hang Li (Huawei Technologies), Piotr Mirowski (Google DeepMind), Alessandro Moschitti (Qatar Computing Research Inst.), Pavel Serdyukov (Yandex), Fabrizio Silvestri (Yahoo Labs), and Alessandro Sordoni (University of Montreal). We would also like to thank the conference organizers for hosting the workshop and accepting our workshop proposal. Finally, we want to thank all the authors, speakers and other participants for making the first workshop on neural information retrieval a success.

References

[1] N. O. Amer, P. Mulhem, and M. Géry. Toward word embedding for personalized information retrieval. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[2] D. Bahdanau, K. Cho, and Y. Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[3] G. Balikas and M.-R. Amini. An empirical study on large scale text classification with skip-gram embeddings. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[4] A. Borisov, I. Markov, M. de Rijke, and P. Serdyukov. A context-aware time model for web search. In *Proc. SIGIR*, pages 205–214. ACM, July 2016.

[5] A. Borisov, I. Markov, M. de Rijke, and P. Serdyukov. A neural click model for web search. In *Proc. WWW*, pages 531–541. ACM, April 2016.

[6] F. Carrara, A. Esuli, T. Fagni, F. Falchi, and A. M. Fernández. Picture it in your mind: Generating high level visual representations from textual descriptions. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[7] D. Cohen, Q. Ai, and W. B. Croft. Adaptability of neural networks on varying granularity ir tasks. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[8] N. Craswell, W. B. Croft, J. Guo, B. Mitra, and M. de Rijke. Neu-IR: The SIGIR 2016 workshop on neural information retrieval. In *Proc. SIGIR*, 2016.

[9] F. Diaz, B. Mitra, and N. Craswell. Query expansion with locally-trained word embeddings. In *Proc. ACL*, 2016.

[10] H. Fang, S. Gupta, F. Iandola, R. Srivastava, L. Deng, P. Dollár, J. Gao, X. He, M. Mitchell, J. Platt, et al. From captions to visual concepts and back. *arXiv preprint arXiv:1411.4952*, 2014.

[11] D. Ganguly, D. Roy, M. Mitra, and G. J. Jones. Word embedding based generalized language model for information retrieval. In *Proc. SIGIR*, pages 795–798. ACM, 2015.

[12] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-r. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, 29(6):82–97, 2012.

[13] P.-S. Huang, X. He, J. Gao, L. Deng, A. Acero, and L. Heck. Learning deep structured semantic models for web search using clickthrough data. In *Proc. CIKM*, pages 2333–2338. ACM, 2013.

[14] R. Jozefowicz, O. Vinyals, M. Schuster, N. Shazeer, and Y. Wu. Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410*, 2016.

[15] N. Kanhabua, H. Ren, and T. B. Moeslund. Learning dynamic classes of events using stacked multilayer perceptron networks. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[16] T. Kenter and M. de Rijke. Short text similarity with word embeddings. In *Proc. CIKM*, pages 1411–1420, 2015.

[17] T. Kenter, A. Borisov, and M. de Rijke. Siamese CBOW: Optimizing word embeddings for sentence representations. In *Proc. ACL*. ACL, August 2016.

[18] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[19] W. D. Lewis. Skype translator: Breaking down language and hearing barriers. In *Translating and the Computer (TC37)*, 2015.

[20] C. Lioma, B. Larsen, C. Petersen, and J. G. Simonsen. Deep learning relevance: Creating relevant information (as opposed to retrieving it). In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[21] P. Luukkonen, M. Koskela, and P. Floréen. Lstm-based predictions for proactive information retrieval. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[22] J. Manotumruksa, C. Macdonald, and I. Ounis. Modelling user preferences using word embeddings for context-aware venue recommendation. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[23] B. Mitra. Exploring session context using distributed representations of queries and reformulations. In *Proc. SIGIR*, pages 3–12. ACM, 2015.

[24] B. Mitra and N. Craswell. Query auto-completion for rare prefixes. In *Proc. CIKM*. ACM, 2015.

[25] B. Mitra, E. Nalisnick, N. Craswell, and R. Caruana. A dual embedding space model for document ranking. *arXiv preprint arXiv:1602.01137*, 2016.

[26] E. Nalisnick, B. Mitra, N. Craswell, and R. Caruana. Improving document ranking with dual word embeddings. In *Proc. WWW*, 2016.

[27] G.-H. Nguyen, L. Tamine, L. Soulier, and N. Bricon-Souf. Toward a deep neural

approach for knowledge-based ir. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[28] L. Pang, Y. Lan, J. Guo, J. Xu, and X. Cheng. A study of matchpyramid models on ad-hoc retrieval. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[29] A. Prakash, C. Brockett, and P. Agrawal. Emulating human conversations using convolutional neural network-based ir. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[30] N. Rekabsaz, M. Lupu, and A. Hanbury. Uncertainty in neural network word embedding: Exploration of threshold for similarity. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[31] D. Roy, D. Ganguly, M. Mitra, and G. J. Jones. Representing documents and queries as sets of word embedded vectors for information retrieval. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[32] D. Roy, D. Paul, M. Mitra, and U. Garain. Using word embeddings for automatic query expansion. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[33] A. Severyn and A. Moschitti. Learning to rank short text pairs with convolutional deep neural networks. In *Proc. SIGIR*, pages 373–382. ACM, 2015.

[34] S. P. Suggu, K. N. Goutham, M. K. Chinnakotla, and M. Shrivastava. Deep feature fusion network for answer quality prediction in community question answering. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[35] C. Van Gysel, M. de Rijke, and E. Kanoulas. Learning latent vector spaces for product search. In *Proc. CIKM*. ACM, October 2016.

[36] C. Van Gysel, M. de Rijke, and M. Worring. Unsupervised, efficient and semantic expertise retrieval. In *Proc. WWW*, pages 1069–1079. ACM, April 2016.

[37] O. Vinyals and Q. Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[38] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. *arXiv preprint arXiv:1411.4555*, 2014.

[39] Y. Wen, W. Zhang, R. Luo, and J. Wang. Learning text representation using recurrent convolutional neural network with highway layers. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[40] J. Yang, J. Tong, R. J. Stones, Z. Zhang, B. Ye, G. Wang, and X. Liu. Selective term proximity scoring via bp-ann. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[41] X. Yang, C. Macdonald, and I. Ounis. Using word embeddings in twitter election classification. In *Neu-IR: The SIGIR 2016 Workshop on Neural Information Retrieval*, 2016.

[42] G. Zheng and J. Callan. Learning to reweight terms with distributed representations. In *Proc. SIGIR*, pages 575–584. ACM, 2015.