# Report on the Second SIGIR Workshop on Neural Information Retrieval (Neu-IR'17)

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

[1]Microsoft [2]University of Massachusetts [3]University of Amsterdam [4]Chinese Academy of Sciences [5]University College London

**Abstract**

The second SIGIR workshop on neural information retrieval (Neu-IR'17) took place on August 11, 2017, in Tokyo, Japan. Following the successful 2016 edition, the workshop continued to serve as a forum for academic and industrial researchers to present new work on neural methods for retrieval. In addition, a special track was organized focusing on resources for evaluation and reproducibility, including proposals for public benchmarking datasets and shared model repositories. A total of 19 papers—which included five special track papers—were presented in the form of oral or poster presentations. Organizers of four of the TREC 2017 tracks were invited to present at the workshop on how these IR tasks may be suitable for evaluating recent data-hungry neural approaches. The full-day workshop—with more than 170 registrants—concluded with an engaging panel discussion.

## 1 Introduction

Following the popularity of the 2016 edition [1, 2], the Neu-IR workshop on neural information retrieval (IR) returned to SIGIR 2017 in Tokyo, Japan. The workshop was co-located with the main conference, where a significant number of papers presented this year focused on neural approaches. On the first day of the conference, SIGIR 2017 also hosted the Neural Networks for Information Retrieval (NN4IR) tutorial [10] which was attended by more than 250 participants which—along with other well-attended tutorials [12, 14] and recent surveys [15, 20]—is further testament to a burgeoning community. In spite of the excitement surrounding the field, however, many important questions that came up at the 2016 workshop still remain relevant a year later.

- **Reproducibility:** Deep neural architectures and neural models that incorporate representation learning for retrieval require large quantities of labeled training examples [16, 17]. Due to the lack of large scale public datasets, many recently published models have been evaluated on private industry datasets and non-standard tasks. The rapid increase in the number of new models in the literature also puts significant burden on future authors with respect to the implementation of baselines. The widespread use of different neural toolkits for model implementation and insufficient documentation of the hyper-parameter choices compound the issue of reproducibility of these new methods.

- **Generalizability:** Many recent neural IR papers report improvements on IR tasks using new architectures. However, it is often less clear how these improvements may generalize to other architectures and methods, or how the same model may perform on other IR tasks. It is important that from these empirical improvements we glean a better understanding of the set of key principles that should guide our design and architecture choices for different types of IR tasks.
- **Interpretability:** Machine learning models, such as deep neural networks, are infamously hard to interpret because of the non-linearity of the functions they learn. The improvements from these *black box* models on IR tasks may come at the cost of reduced focus on gaining IR insights which may limit our progress in the long term. Unless we emphasize on better understanding of the connection between these new models and classical IR approaches, we also run the risk of wasting time and effort re-discovering many existing IR principles.
- **Evaluation:** More work is necessary in developing good metrics and benchmarks for emerging IR scenarios, such as proactive retrieval and conversational IR. A combined progress on metrics and models is likely key to making new breakthroughs on these emerging IR tasks.

This year's workshop continued to serve as an important forum for the community to voice and discuss these challenges. A special track was announced to specifically solicit proposals that address the challenges to reproducibility and evaluation. In addition, invited talks from TREC track organizers and a panel discussion were organized to further encourage discussions on these topics. The workshop received a strong response—with 25 submissions and more than 170 registrations—which indicate that in spite of the strong presence of neural IR papers at the main conference, the workshop continues to play an important role in bringing the community together to discuss the future directions for the field.

# 2 Scope and format

This year the workshop organizers solicited [3] submissions to two separate tracks. The focus of the *special track* was on tackling challenges around training, evaluation and the reproducibility of deep neural network models for IR. In particular, it invited proposals for,

- New large scale benchmark collections appropriate for training and evaluating deep neural network models with millions (or billions) of parameters.
- Building a central shared model repository without enforcing the use of any specific neural network toolkit.
- Making appropriate hardware resources (e.g., GPUs) available for academic research.
- New tools and bindings to enable smooth interfacing between traditional IR frameworks and recent neural network toolkits.
- Standardizing frameworks appropriate for evaluating deep neural network models.
- Automatic and semi-automatic methods for generating training material at scale.

In addition, the *general track* 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 IR, entity ranking
  - Learning to rank combined with neural network based representation learning
  - User and task modelling, personalized search, 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
  - Real-world use cases, deployment at scale

All papers were peer reviewed (single-blind) by the program committee and judged by their relevance to the workshop, either to the special topic or to the general themes identified above, and their potential to generate discussions. Papers were limited to two to eight pages in length. Similar to last year, all submissions were considered non-archival, and were required to be uploaded to `https://arXiv.org` if accepted for presentation at the workshop.

# 3 Keynote

The keynote talk at the workshop was given by Yelong Shen from the Deep Learning Technology Center, Microsoft Research. The talk, titled *Deep Neural Networks for IR - from Web Search to Reading Comprehension*, provided a brief overview of the Convolutional DSSM [25] for short text matching. The latter part of the talk focused on the machine reading and comprehension task, and the recently proposed ReasoNet model [26]. The keynote concluded with a discussion on symbolic neural nets and knowledge representation.

# 4 Accepted papers

A total of 19 papers were accepted for presentation at the workshop, including five from the special track [4, 5, 13, 27, 28] and 14 from the general track [6–9, 11, 18, 19, 21–24, 29–31]. All accepted papers were presented as posters, and six papers—two from the special track

Table 1: Speakers and topics for the "TREC talks" session.

| TREC track | Speaker | Website |
|---|---|---|
| Complex Answers Retrieval | Laura Dietz | http://trec-car.cs.unh.edu/ |
| Real-Time Summarization | Jimmy Lin | http://trecrts.github.io/ |
| Dynamic Domain | Grace Hui Yang | http://trec-dd.org/ |
| Open Search | Krisztian Balog | http://trec-open-search.org/ |

and four from the general track—were additionally selected for oral presentations. The full list of accepted papers is available on the workshop website.[1]

# 5 TREC talks

Organizers of four TREC 2017 tracks were invited to come present at the workshop. The focus of these presentations was on the suitability of the respective IR tasks for evaluating data-hungry neural models. Laura Dietz presented early benchmarking results on the Complex Answer Retrieval track [17] with some promising improvements from deep neural networks. Table 1 lists all the TREC tracks covered, and the respective speakers, from the session.

# 6 Panel discussion

The workshop concluded with a panel discussion chaired by Maarten de Rijke with Fernando Diaz, Claudia Hauff, and Jian-Yun Nie as the invited panelists. Some of the common themes re-emerged during the panel discussion.

- What does success look like for neural IR? How do we make sure we are making real progress with these new machine learning approaches?
- How do we overcome the issue of lack of public datasets for benchmarking data-hungry machine learning methods? As we identify promising sources of synthetic or real training data for neural information retrieval, should we also make an effort to make the data more readily available to non-IR experts? Such pre-packaged learning to rank datasets allow ML-oriented researchers to participate without needing to download the corpus, index it, or extract features.
- What are the IR problems that stand to gain most from advances in neural networks?
- Should neural IR research focus on classic IR tasks such as ad-hoc retrieval, or should there be a stronger emphasis on new scenarios such as conversational IR?
- What are downsides of focusing too much on neural IR? What are the convincing argument for why new phd students should not work on neural IR?

As one of the concluding remarks, Claudia Hauff emphasized the importance of not neglecting IR fundamentals in the pursuit of new approaches.

[1] https://neu-ir.weebly.com/accepted-papers.html

# 7 Conclusions

Neural IR has already gained mainstream popularity within the IR community. At the end of the workshop, a question was posed to the audience on whether there was still a need for a separate workshop on this topic to which there was a strong consensus among those present that such a platform for pointed conversations about the future of the field was indeed useful. Many challenges related to reproducibility and generalizability still remains that solicits deliberate discussions and debates among the community. Hopefully, future editions of this workshop will continue to provide that forum.

# 8 Acknowledgements

The workshop organizers are sincerely grateful to all the program committee members – Alexey Borisov (University of Amsterdam), Mostafa Dehghani (University of Amsterdam), Zhicheng Dou (Renmin University of China), Carsten Eickhoff (ETH Zurich), Christina Lioma (University of Copenhagen), Yiqun Liu (Tsinghua University), Rishabh Mehrotra (University College London), Piotr Mirowski (DeepMind), Alessandro Moschitti (Qatar Computing Research Institute), Pavel Serdyukov (Yandex), Aliaksei Severyn (Google), Fabrizio Silvestri (Facebook), Christophe Van Gysel (University of Amsterdam), Manisha Verma (University College London), Hamed Zamani (University of Massachusetts Amherst), Peng Zhang (Tianjin University), and Qi Zhang (Fudan University). 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 second workshop on neural information retrieval a success.

# References

[1] 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*. ACM, 2016.

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

[3] N. Craswell, W. B. Croft, M. de Rijke, J. Guo, and B. Mitra. Neu-ir'17: Neural information retrieval. In *Proc. SIGIR*. ACM, 2017.

[4] M. Dehghani, H. Azarbonyad, J. Kamps, and M. de Rijke. Share your model instead of your data: Privacy preserving mimic learning for ranking. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[5] Y. Fan, L. Pang, J. Hou, J. Guo, Y. Lan, and X. Cheng. Matchzoo: A toolkit for deep text matching. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[6] U. Gupta, A. Chatterjee, R. Srikanth, and P. Agrawal. A sentiment-and-semantics-based approach for emotion detection in textual conversations. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[7] K. Hui, A. Yates, K. Berberich, and G. de Melo. Re-pacrr: A context and density-aware neural information retrieval model. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[8] R. Jagerman, J. Kiseleva, and M. de Rijke. Modeling label ambiguity for neural list-wise learning to rank. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[9] A. Kale, T. Taula, S. Hewavitharana, and A. Srivastava. Towards semantic query segmentation. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[10] T. Kenter, A. Borisov, C. Van Gysel, M. Dehghani, M. de Rijke, and B. Mitra. Neural networks for information retrieval (nn4ir). In *Proc. SIGIR*. ACM, 2017.

[11] P. Khosla, M. Basu, K. Ghosh, and S. Ghosh. Microblog retrieval for post-disaster relief: Applying and comparing neural ir models. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[12] H. Li and Z. Lu. Deep learning for information retrieval. In *Proc. SIGIR*. ACM, 2016.

[13] S. MacAvaney, K. Hui, and A. Yates. An approach for weakly-supervised deep information retrieval. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[14] B. Mitra and N. Craswell. Neural text embeddings for information retrieval. In *Proc. WSDM*, pages 813–814. ACM, 2017.

[15] B. Mitra and N. Craswell. Neural models for information retrieval. *arXiv preprint arXiv:1705.01509*, 2017.

[16] B. Mitra, F. Diaz, and N. Craswell. Learning to match using local and distributed representations of text for web search. In *Proc. WWW*, pages 1291–1299, 2017.

[17] F. Nanni, B. Mitra, M. Magnusson, and L. Dietz. Benchmark for complex answer retrieval. In *Proc. ICTIR*. ACM, 2017.

[18] D. T. Nguyen, S. Joty, B. E. A. Boussaha, and M. de Rijke. Thread reconstruction in conversational data using neural coherence models. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[19] R. Nogueira and K. Cho. Task-oriented query reformulation with reinforcement learning. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[20] K. D. Onal, Y. Zhang, I. S. Altingovde, M. M. Rahman, et al. Neural information retrieval: At the end of the early years. *Information Retrieval Journal*, 2017.

[21] L. Pang, Y. Lan, J. Guo, J. Xu, and X. Cheng. A deep investigation of deep ir models. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[22] J. Rao, H. He, H. Zhang, F. Ture, R. Sequiera, S. Mohammed, and J. Lin. Integrating lexical and temporal signals in neural ranking models for searching social media streams. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[23] N. Rekabsaz, B. Mitra, M. Lupu, and A. Hanbury. Toward incorporation of relevant documents in word2vec. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural*

*Information Retrieval*, 2017.

[24] R. Sequiera, G. Baruah, Z. Tu, S. Mohammed, J. Rao, H. Zhang, and J. Lin. Exploring the effectiveness of convolutional neural networks for answer selection in end-to-end question answering. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[25] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil. Learning semantic representations using convolutional neural networks for web search. In *Proc. WWW*, pages 373–374, 2014.

[26] Y. Shen, P.-S. Huang, J. Gao, and W. Chen. Reasonet: Learning to stop reading in machine comprehension. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1047–1055. ACM, 2017.

[27] Z. Tu, M. Crane, R. Sequiera, J. Zhang, and J. Lin. An exploration of approaches to integrating neural reranking models in multi-stage ranking architectures. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[28] C. Van Gysel, M. de Rijke, and E. Kanoulas. Semantic entity retrieval toolkit. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[29] C. Wu, M. Yan, and L. Si. Session-aware information embedding for e-commerce product recommendation. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[30] L. Yang, H. Zamani, Y. Zhang, J. Guo, and W. B. Croft. Neural matching models for question retrieval and next question prediction in conversation. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.

[31] A. Yates and K. Hui. De-pacrr: Exploring layers inside the pacrr model. In *Neu-IR'17: The SIGIR 2017 Workshop on Neural Information Retrieval*, 2017.