

References I

- Qingyao Ai, Liu Yang, Jiafeng Guo, and W. Bruce Croft. 2016a. Analysis of the Paragraph Vector Model for Information Retrieval. In *ICTIR*. ACM, 133–142.
- Qingyao Ai, Liu Yang, Jiafeng Guo, and W Bruce Croft. 2016b. Improving language estimation with the paragraph vector model for ad-hoc retrieval. In *SIGIR*. ACM, 869–872.
- Qingyao Ai, Yongfeng Zhang, Keping Bi, Xu Chen, and Bruce W. Croft. 2017. Learning a Hierarchical Embedding Model for Personalized Product Search. In *SIGIR*.
- Nima Asadi, Donald Metzler, Tamer Elsayed, and Jimmy Lin. 2011. Pseudo test collections for learning web search ranking functions. In *SIGIR*. ACM, 1073–1082.
- Leif Azzopardi, Maarten de Rijke, and Krisztian Balog. 2007. Building simulated queries for known-item topics: An analysis using six European languages. In *SIGIR*. ACM.
- Marco Baroni, Georgiana Dinu, and Germán Kruszewski. 2014. Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors.. In *ACL (1)*. 238–247.
- Steven M Beitzel, Eric C Jensen, Abdur Chowdhury, and David Grossman. 2003. Using titles and category names from editor-driven taxonomies for automatic evaluation. In *CIKM*. ACM, 17–23.
- Yoshua Bengio and Jean-Sébastien Senécal. 2008. Adaptive importance sampling to accelerate training of a neural probabilistic language model. *IEEE Transactions on Neural Networks* 19, 4 (2008), 713–722.
- Yoshua Bengio, Jean-Sébastien Senécal, and others. 2003. Quick Training of Probabilistic Neural Nets by Importance Sampling.. In *AISTATS*.
- Richard Berendsen, Manos Tsagkias, Wouter Weerkamp, and Maarten de Rijke. 2013. Pseudo test collections for training and tuning microblog rankers. In *SIGIR*. ACM, 53–62.
- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *JMLR* 3 (2003), 993–1022.
- Antoine Bordes and Jason Weston. 2017. Learning end-to-end goal-oriented dialog. *ICLR* (2017).
- Alexey Borisov, Ilya Markov, Maarten de Rijke, and Pavel Serdyukov. 2016. A neural click model for web search. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 531–541.
- Chris Burges. 2015. RankNet: A ranking retrospective. (2015).
<https://www.microsoft.com/en-us/research/blog/ranknet-a-ranking-retrospective/> Accessed July 16, 2017.

References II

- Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. 2005. Learning to rank using gradient descent. In *Proceedings of the 22nd international conference on Machine learning*. ACM, 89–96.
- Christopher JC Burges. 2010. From ranknet to lambdarank to lambdamart: An overview. *Learning* 11, 23-581 (2010), 81.
- Christopher JC Burges, Robert Ragno, and Quoc Viet Le. 2006. Learning to rank with nonsmooth cost functions. In *NIPS*, Vol. 6. 193–200.
- Zhe Cao, Tao Qin, Tie-Yan Liu, Ming-Feng Tsai, and Hang Li. 2007. Learning to rank: from pairwise approach to listwise approach. In *Proceedings of the 24th international conference on Machine learning*. ACM, 129–136.
- Gabriele Capannini, Claudio Lucchese, Franco Maria Nardini, Salvatore Orlando, Raffaele Perego, and Nicola Tonellotto. 2016. Quality versus efficiency in document scoring with learning-to-rank models. *IPM* 52, 6 (2016), 1161–1177.
- Ben Carterette and Rosie Jones. 2008. Evaluating search engines by modeling the relationship between relevance and clicks. In *NIPS*. 217–224.
- Wei Chen, Tie-Yan Liu, Yanyan Lan, Zhi-Ming Ma, and Hang Li. 2009. Ranking measures and loss functions in learning to rank. In *Advances in Neural Information Processing Systems*. 315–323.
- Charles LA Clarke, Maheedhar Kolla, Gordon V Cormack, Olga Vechtomova, Azin Ashkan, Stefan Büttcher, and Ian MacKinnon. 2008. Novelty and diversity in information retrieval evaluation. In *SIGIR*. ACM, 659–666.
- David Cossock and Tong Zhang. 2006. Subset ranking using regression. In *COLT*, Vol. 6. Springer, 605–619.
- Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the American Society for Information Science* 41, 6 (1990), 391–407.
- Mostafa Dehghani, Hamed Zamani, Aliaksei Severyn, Jaap Kamps, and W Bruce Croft. 2017. Neural Ranking Models with Weak Supervision. In *SIGIR*.
- Bhuwan Dhingra, Lihong Li, Xiujuan Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmed, and Li Deng. 2017. Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*.
- Fernando Diaz, Bhaskar Mitra, and Nick Craswell. 2016. Query expansion with locally-trained word embeddings. In *ACL*.
- John Rupert Firth. 1957. *Papers in Linguistics 1934-1951*. Oxford University Press.
- Yoav Freund, Raj Iyer, Robert E Schapire, and Yoram Singer. 2003. An efficient boosting algorithm for combining preferences. *Journal of machine learning research* 4, Nov (2003), 933–969.

References III

- Norbert Fuhr. 1989. Optimum polynomial retrieval functions based on the probability ranking principle. *ACM Transactions on Information Systems (TOIS)* 7, 3 (1989), 183–204.
- Debasis Ganguly, Dwaipayan Roy, Mandar Mitra, and Gareth JF Jones. 2015. Word embedding based generalized language model for information retrieval. In *SIGIR*. ACM, 795–798.
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep learning*. MIT press.
- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.
- Joshua Goodman. 2001. Classes for fast maximum entropy training. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, Vol. 1. IEEE, 561–564.
- Artem Grotov and Maarten de Rijke. 2016. Online Learning to Rank for Information Retrieval: SIGIR 2016 Tutorial. In *SIGIR*. ACM, 1215–1218.
- Jiafeng Guo, Yixing Fan, Qingyao Ai, and W Bruce Croft. 2016. A Deep Relevance Matching Model for Ad-hoc Retrieval. ACM, 55–64.
- Michael Gutmann and Aapo Hyvärinen. 2010. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models.. In *AISTATS*, Vol. 1. 6.
- Zellig S Harris. 1954. Distributional structure. *Word* 10, 2-3 (1954), 146–162.
- Ralf Herbrich, Thore Graepel, and Klaus Obermayer. 2000. Large margin rank boundaries for ordinal regression. (2000).
- Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. In *NIPS*.
- Daniel Hewlett, Alexandre Lacoste, Llion Jones, Illia Polosukhin, Andrew Fandrianto, Jay Han, Matthew Kelcey, and David Berthelot. 2016. WIKIREADING: A Novel Large-scale Language Understanding Task over Wikipedia. In *ACL*.
- S. Hochreiter and J. Schmidhuber. 1997. Long short-term memory. *Neural Computation* 9, 8 (1997), 1735–1780.
- Thomas Hofmann. 1999. Probabilistic latent semantic indexing. In *SIGIR*. ACM, 50–57.
- 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*. 2042–2050.

References IV

- Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. 2013. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*. ACM, 2333–2338.
- Bouke Huurnink, Katja Hofmann, and Maarten de Rijke. 2010a. Simulating searches from transaction logs. *Simulation of Interaction* (2010), 21.
- Bouke Huurnink, Katja Hofmann, Maarten De Rijke, and Marc Bron. 2010b. Validating query simulators: An experiment using commercial searches and purchases. In *CLEF*. Springer, 40–51.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2016. Image-to-Image Translation with Conditional Adversarial Networks. *arxiv* (2016).
- Rolf Jagerman, Julia Kiseleva, and Maarten de Rijke. 2017. Modeling Label Ambiguity for Neural List-Wise Learning to Rank. In *Neu-IR SIGIR Workshop*.
- Kalervo Järvelin and Jaana Kekäläinen. 2000. IR Evaluation Methods for Retrieving Highly Relevant Documents. In *SIGIR*. ACM, 41–48.
- Sébastien Jean, Kyunghyun Cho, Roland Memisevic, and Yoshua Bengio. 2014. On Using Very Large Target Vocabulary for Neural Machine Translation. *arXiv preprint arXiv:1412.2007* (2014).
- Shihao Ji, SVN Vishwanathan, Nadathur Satish, Michael J Anderson, and Pradeep Dubey. 2015. Blackout: Speeding up recurrent neural network language models with very large vocabularies. *arXiv preprint arXiv:1511.06909* (2015).
- Thorsten Joachims. 2003. Evaluating Retrieval Performance Using Clickthrough Data. In *Text Mining*. 79–96.
- Thorsten Joachims. 2006. Training linear SVMs in linear time. In *KDD*. ACM, 217–226.
- Thorsten Joachims, Dayne Freitag, and Tom Mitchell. 1997. Webwatcher: A tour guide for the world wide web. In *IJCAI*. 770–777.
- Rafal Jozefowicz, Oriol Vinyals, Mike Schuster, Noam Shazeer, and Yonghui Wu. 2016. Exploring the limits of language modeling. *arXiv preprint arXiv:1602.02410* (2016).
- Tom Kenter, Alexey Borisov, and Maarten de Rijke. 2016. Siamese cbow: Optimizing word embeddings for sentence representations. *arXiv preprint arXiv:1606.04640* (2016).
- Tom Kenter and Maarten de Rijke. 2017. Attentive Memory Networks: Efficient Machine Reading for Conversational Search. In *Proceedings of the The First International Workshop on Conversational Approaches to Information Retrieval (CAIR'17)*.
- Jinyoung Kim and W. Bruce Croft. 2009. Retrieval Experiments Using Pseudo-desktop Collections. In *CIKM*. ACM, 1297–1306.
- Quoc V Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *ICML*. 1188–1196.

References V

- Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A diversity-promoting objective function for neural conversation models. In *NAACL-HLT 2016*. 110–119.
- Jiwei Li, Michel Galley, Chris Brockett, Georgios P. Spithourakis, Jianfeng Gao, and William B. Dolan. 2016. A Persona-Based Neural Conversation Model. In *ACL 2016*.
- Jiwei Li, Will Monroe, Tianlin Shi, Alan Ritter, and Dan Jurafsky. 2017. Adversarial learning for neural dialogue generation. *arXiv preprint arXiv:1701.06547* (2017).
- Ping Li, Qiang Wu, and Christopher J Burges. 2008. Mcrank: Learning to rank using multiple classification and gradient boosting. In *Advances in neural information processing systems*. 897–904.
- Chia-Wei Liu, Ryan Lowe, Iulian V Serban, Michael Noseworthy, Laurent Charlin, and Joelle Pineau. 2016. How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *EMNLP 2016* (2016).
- Qiang Liu, Feng Yu, Shu Wu, and Liang Wang. 2015. A convolutional click prediction model. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 1743–1746.
- Tie-Yan Liu. 2009. Learning to rank for information retrieval. *Foundations and Trends® in Information Retrieval* 3, 3 (2009), 225–331.
- Ryan Lowe, Nissan Pow, Iulian Serban, and Joelle Pineau. 2015. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *SIGDIAL 2015*. 285–294.
- R Duncan Luce. 2005. *Individual choice behavior: A theoretical analysis*. Courier Corporation.
- S. MacAvaney, K. Hui, and A. Yates. 2017. An Approach for Weakly-Supervised Deep Information Retrieval. *arXiv 1707.00189*. (2017).
- Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. 3111–3119.
- Alexander Miller, Adam Fisch, Jesse Dodge, Amir-Hossein Karimi, Antoine Bordes, and Jason Weston. 2016. Key-Value Memory Networks for Directly Reading Documents. In *EMNLP*.
- Bhaskar Mitra. 2015. Exploring session context using distributed representations of queries and reformulations. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 3–12.
- Bhaskar Mitra and Nick Craswell. 2017. An Introduction to Neural Information R. *Foundations and Trends® in Information Retrieval (to appear)* (2017).

References VI

- Bhaskar Mitra, Fernando Diaz, and Nick Craswell. 2017. Learning to Match Using Local and Distributed Representations of Text for Web Search. 1291–1299.
- Andriy Mnih and Geoffrey E Hinton. 2009. A scalable hierarchical distributed language model. In *Advances in neural information processing systems*. 1081–1088.
- Andriy Mnih and Yee Whye Teh. 2012. A fast and simple algorithm for training neural probabilistic language models. *arXiv preprint arXiv:1206.6426* (2012).
- Frederic Morin and Yoshua Bengio. 2005. Hierarchical Probabilistic Neural Network Language Model.. In *Aistats*, Vol. 5. Citeseer, 246–252.
- Eric Nalisnick, Bhaskar Mitra, Nick Craswell, and Rich Caruana. 2016. Improving Document Ranking with Dual Word Embeddings.
- Federico Nanni, Bhaskar Mitra, Matt Magnusson, and Laura Dietz. 2017. Benchmark for Complex Answer Retrieval. ACM.
- Liang Pang, Yanyan Lan, Jiafeng Guo, Jun Xu, Shengxian Wan, and Xueqi Cheng. 2016. Text Matching as Image Recognition.. In *AAAI*. 2793–2799.
- Filip Radlinski and Thorsten Joachims. 2005. Query chains: learning to rank from implicit feedback. In *KDD*. ACM, 239–248.
- Filip Radlinski, Madhu Kurup, and Thorsten Joachims. 2008. How does clickthrough data reflect retrieval quality?. In *CIKM*. ACM, 43–52.
- Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. <http://is.muni.cz/publication/884893/en>.
- Dwaipayan Roy, Debjyoti Paul, Mandar Mitra, and Utpal Garain. 2016. Using Word Embeddings for Automatic Query Expansion. *arXiv preprint arXiv:1606.07608* (2016).
- Ruslan Salakhutdinov and Geoffrey Hinton. 2009. Semantic hashing. *Int. J. Approximate Reasoning* 50, 7 (2009), 969–978.
- D. Sculley. 2009. Large scale learning to rank. In *In NIPS 2009 Workshop on Advances in Ranking*.
- Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C Courville, and Joelle Pineau. 2016. Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models.. In *AAAI*. 3776–3784.
- Pararth Shah, Dilek Hakkani-Tür, and Larry Heck. 2016. Interactive reinforcement learning for task-oriented dialogue management. In *NIPS 2016 Deep Learning for Action and Interaction Workshop*.
- Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *ACL 2016*.

References VII

- Yelong Shen, Xiaodong He, Jianfeng Gao, Li Deng, and Grégoire Mesnil. 2014. A latent semantic model with convolutional-pooling structure for information retrieval. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*. ACM, 101–110.
- Alessandro Sordoni, Yoshua Bengio, Hossein Vahabi, Christina Lioma, Jakob G. Simonson, and Jian-Yun Nie. 2015a. A Hierarchical Recurrent Encoder-Decoder for Generative Context-Aware Query Suggestion. In *CIKM 2015*.
- Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Meg Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015b. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. In *Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT)*.
- K. Sparck Jones and C.J. van Rijsbergen. 1976. *Report on the need for and provision of an 'ideal' information retrieval test collection*. Technical Report. Computer Laboratory, Cambridge University.
- Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. 2015. End-To-End Memory Networks. In *NIPS*.
- Jean Tague, Michael Nelson, and Harry Wu. 1980. Problems in the simulation of bibliographic retrieval systems. In *SIGIR*. Butterworth & Co., 236–255.
- Jean Tague and Michael J Nelson. 1981. Simulation of user judgments in bibliographic retrieval systems. In *ACM SIGIR Forum*, Vol. 16. ACM, 66–71.
- Jörg Tiedemann. 2009. News from OPUS-A collection of multilingual parallel corpora with tools and interfaces. In *Recent advances in natural language processing*, Vol. 5. 237–248.
- Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. 2016. Learning Latent Vector Spaces for Product Search. In *CIKM*. ACM, 165–174.
- Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. 2017a. Semantic Entity Retrieval Toolkit. In *Neu-IR SIGIR Workshop*.
- Christophe Van Gysel, Maarten de Rijke, and Evangelos Kanoulas. 2017b. Structural Regularities in Expert Vector Spaces. In *ICTIR*. ACM.
- Christophe Van Gysel, Maarten de Rijke, and Marcel Worring. 2016. Unsupervised, Efficient and Semantic Expertise Retrieval. In *WWW*. ACM, 1069–1079.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. *arXiv* (2017).

References VIII

- Ashish Vaswani, Yingdong Zhao, Victoria Fossum, and David Chiang. 2013. Decoding with Large-Scale Neural Language Models Improves Translation.. In *EMNLP*. 1387–1392.
- Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *ICML Deep Learning Workshop 2015* (2015).
- Ivan Vulić and Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. In *SIGIR*. ACM, 363–372.
- Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. 2017. A network-based end-to-end trainable task-oriented dialogue system. In *EACL*.
- Jason Weston, Sumit Chopra, and Antoine Bordes. 2015. Memory networks. In *ICLR*.
- Fen Xia, Tie-Yan Liu, Jue Wang, Wensheng Zhang, and Hang Li. 2008. Listwise approach to learning to rank: theory and algorithm. In *Proceedings of the 25th international conference on Machine learning*. ACM, 1192–1199.
- Emine Yilmaz and Stephen Robertson. 2009. Deep Versus Shallow Judgments in Learning to Rank. In *SIGIR*. ACM, 662–663.
- Hamed Zamani and W. Bruce Croft. 2016a. Embedding-based Query Language Models. In *ICTIR*. ACM, 147–156.
- Hamed Zamani and W. Bruce Croft. 2016b. Estimating Embedding Vectors for Queries. In *ICTIR*. ACM, 123–132.
- Hamed Zamani and W Bruce Croft. 2017. Relevance-based Word Embedding. In *SIGIR*.
- Yuyu Zhang, Hanjun Dai, Chang Xu, Jun Feng, Taifeng Wang, Jiang Bian, Bin Wang, and Tie-Yan Liu. 2014. Sequential Click Prediction for Sponsored Search with Recurrent Neural Networks.. In *AAAI*. 1369–1375.
- Guoqing Zheng and Jamie Callan. 2015. Learning to reweight terms with distributed representations. In *SIGIR*. ACM, 575–584.
- Zhaohui Zheng, Keke Chen, Gordon Sun, and Hongyuan Zha. 2007. A regression framework for learning ranking functions using relative relevance judgments. In *SIGIR*. ACM, 287–294.
- Guido Zuccon, Bevan Koopman, Peter Bruza, and Leif Azzopardi. 2015. Integrating and Evaluating Neural Word Embeddings in Information Retrieval. In *20th Australasian Document Computing Symposium*. ACM, Article 12, 8 pages.

Notation

Meaning	Notation
Single query	q
Single document	d
Set of queries	Q
Collection of documents	D
Term in query q	t_q
Term in document d	t_d
Full vocabulary of all terms	T
Set of ranked results retrieved for query q	R_q
Result tuple (document d at rank i)	$\langle i, d \rangle$, where $\langle i, d \rangle \in R_q$
Relevance label of document d for query q	$rel_q(d)$
d_i is more relevant than d_j for query q	$rel_q(d_i) > rel_q(d_j)$, or $d_i \succ_q d_j$
Frequency of term t in document d	$tf(t, d)$
Number of documents that contain term t	$df(t)$
Vector representation of text z	\vec{z}
Probability function for an event \mathcal{E}	$p(\mathcal{E})$
\mathbb{R}	The set of real numbers

Acknowledgments



All content represents the opinion of the author(s), which is not necessarily shared or endorsed by their employers and/or sponsors.