Inf Retrieval (2010) 13:197–200
DOI 10.1007/s10791-009-9120-1

Introduction to special issue on learning to rank for information retrieval

Tie-Yan Liu · Thorsten Joachims · Hang Li · Chengxiang Zhai

Received: 16 November 2009 / Accepted: 16 November 2009 / Published online: 20 December 2009

Learning to rank has emerged as an active and growing area of research both in information retrieval (IR) and machine learning (ML). Many IR problems are by nature ranking problems, and many IR technologies can be potentially enhanced by using learning to rank techniques. These include document retrieval, definition ranking (Xu et al. 2005), question answering (Banerjee et al. 2009; Surdeanu et al. 2008; Verberne et al. 2009), multimedia retrieval (Yang et al. 2008a; Yang et al. 2008b), text summarization (Metzler et al. 2008) and advertisement (Ciaramita et al. 2008; Mao 2009).

Taking document retrieval as example, the learning to rank process can be described as follows. In the training phase, each query-document pair is represented by a feature vector, and the relevance between the document and the query is given as ground truth. The goal of learning to rank is to automatically learn a ranking model from the training data, such that the model can accurately rank documents by relevance, also for new queries.

In the past years, significant progress has been made on the learning to rank problem and the development of methods.

– Several major approaches have been proposed, with different loss functions and formulations. The pointwise approach (Cossock et al. 2006; Crammer et al. 2002; Li et al. 2007) regards ranking as a regression or classification problem, and views single documents as learning instances; the pairwise approach (Burges et al. 2005; Freund et al. 2003; Joachims 2002) formulates ranking as a pairwise classification problem,

T.-Y. Liu (✉) · H. Li
Microsoft Research Asia, Beijing, China
e-mail: Tie-Yan.Liu@microsoft.com

H. Li
e-mail: hangli@microsoft.com

T. Joachims
Cornell University, Ithaca, NY, USA
e-mail: tj@cs.cornell.edu

C. Zhai
University of Illinois at Urbana-Champaign, Champaign, IL, USA
e-mail: czhai@cs.uiuc.edu

and regards document pairs as learning instances; the listwise approach (Cao et al. 2007; Talyor et al. 2008; Xia et al. 2008; Yue et al. 2007) defines the loss functions on all the documents associated with a query, in a measure-specific manner (also referred to as direct optimization of IR measures) or non-measure specific manner.
– Benchmark datasets like LETOR (Liu et al. 2007) have been released to facilitate the research on learning to rank.[1] Many research papers on learning to rank have used these datasets for their experiments, which makes their results easy to compare.
– Several activities regarding learning to rank have been organized at major conferences. For example, the workshop series on Learning to Rank for IR (2007–2009), the workshop on Beyond Binary Relevance: Preferences, Diversity, and Set-Level Judgments (2008), and the workshop on Redundancy, Diversity, and Interdependent Document Relevance (2009) have been organized at SIGIR. Active participation of the researchers in these activities has demonstrated the continued interests from the research community on the topic of learning to rank.
– Learning to rank has also become a key technology in the industry. Several major search engine companies are using various learning to rank technologies to train their ranking models.[2]

Given the increasing impact of learning to rank in both the research community and the industry, we realized the necessity of organizing a special issue of the Journal of Information Retrieval to discuss the topic. A call for paper was issued on March 6th, 2009. Twenty five papers were received by April 24th, 2009, and each was reviewed by at least three reviewers. Six papers were selected for the special issue based on their qualities. The papers provide different perspectives on the subject and outline the directions for future research. In particular, they try to answer the following important questions.

– Can one develop more efficient and effective algorithms under the framework of major approaches to learning to rank?
– Can ranking models learned from old data be successfully transferred to new data, or from one domain to another domain?
– Are all IR measures suitable for direct optimization?
– Instead of learning from features, can we learn equally effective ranking models from raw content of the query and documents?

Chapelle and Keerthi discussed how to improve the efficiency of Ranking SVMs (Joachims 2002) in their paper entitled "*Efficient Algorithms for Ranking with SVMs*." So far SVMLight has been the only publicly available software for Ranking SVMs.[3] It is slow and, due to incomplete training with it, previous evaluations show Ranking SVMs to have inferior ranking performance. Chapelle and Keerthi proposed new methods based on the primal Newton method to speed up Ranking SVMs training and show that they are five orders of magnitude faster than SVMLight. Evaluation on the LETOR benchmark datasets after complete training using the methods shows that the performance of Ranking SVMs is excellent.

In the paper entitled "*Gradient Descent Optimization of Smoothed Information Retrieval Metrics*," Chapelle and Wu proposed a new method to directly optimize IR

[1] http://research.microsoft.com/~LETOR/

[2] http://glinden.blogspot.com/2005/06/msn-search-and-learning-to-rank.html, http://ysearchblog.com/2008/07/09/boss-the-next-step-in-our-open-search-ecosystem/

[3] http://svmlight.joachims.org/

measures. The basic idea is to minimize a smooth approximation of these measures with gradient descent. Crucial to this kind of approach is the choice of the smoothing factor. Various theoretical analyses were conducted on that choice and an annealing algorithm was proposed to iteratively minimize a less and less smoothed approximation of the measure of interest. Results on the LETOR benchmark datasets show that the proposed algorithm achieves state-of-the-art performances.

Chen et al. argued in their paper entitled "*Knowledge Transfer for Cross Domain Learning to Rank* " that many learning-to-rank applications exist in environments that are subject to rapid change. Training data that was collected in the past can quickly become outdated, since the underlying distribution generating the data changes. Or one may have training data coming from a related, but different domain. For this cross-domain learning to rank problem, they propose two methods based on Ranking SVMs that perform inductive transfer on the feature level and the instance level, respectively.

Wu et al. proposed a new learning-to-rank method that extends LambdaRank to using boosted regression trees as the underlying function class, in the paper entitled "*Ranking, Boosting, and Model Adaptation*." The goal is to maintain LambdaRank's ability to approximately optimize IR measures directly, while improving speed at training and test time. Wu et al. showed that this learning algorithm can naturally exploit existing data from a related task when learning a new task with only little training data. As a special case of the learning-to-rank problem, the paper also gives an algorithm for combining two rankers to optimize a given IR measure.

In the paper "*On the Choice of Effectiveness Measures for Learning to Rank*," Yilmaz and Robertson questioned the current assumption that it is always desirable to directly optimize the target IR measure in learning to rank. They argued that some IR measures may be more "informative"than others for learning and suggested that even if user satisfaction can be measured by an IR measure X, optimizing a search engine based on a different but more informative IR measure Y may actually result in better test performance.

In the paper "*Learning to Rank with (a Lot of) Word Features*," Bai et al. studied learning to rank from a new perspective. Instead of combining different hand-crafted features, they studied how to handle word correlations in the framework of learning to rank. They proposed a supervised semantic indexing algorithm to learn correlations between a query word and a document word. The correlations can be directly used to estimate ranking scores between a query and a document. The authors also proposed methods to improve the efficiency of the algorithm.

The above six papers have solved some existing problems in learning to rank, but have also opened new windows to future research. We hope that this special issue can motivate more researchers in IR and ML to work on learning to rank, and further advance the state of the art.

Finally, we would like to thank the editors of Journal of Information Retrieval for their support in producing this special issue and the reviewers who gave their time and expertises in reading the submitted articles: Alekh Agarwal, Krisztian Balog, Jiang Bian, Chris Burges, Deng Cai, Yunbo Cao, Soumen Chakrabarti, Olivier Chapelle, Zheng Chen, Kevin Duh, Hui Fang, Jun'ichi Fukumoto, Jianfeng Gao, Marcos Andre Goncalves, John Guiver, Paul Heymann, Andreas Hotho, Eyke Huellermeier, Rong Jin, Kristian Kersting, Alexandre Klementiev, Quoc Le, Ping Li, Qiaozhu Mei, Donald Metzler, Ramesh Nallapati, Jian Pei, Tao Qin, Filip Radlinski, Xuehua Shen, Milad Shokouchi, Yang Song, Mike Taylor, Choon Hui Teo, Hamed Valizadegan, Xuanhui Wang, Fen Xia, Jun Xu, Jingfang Xu, Gui-Rong Xue, Kai Yu, Chun-Nam Yu, Yisong Yue, and Zhaohui Zheng.

References

Banerjee, S., Chakrabarti, S., & Ramakrishnan, G. (2009). Learning to rank for quantity consensus queries. In *SIGIR 2009* (pp. 243–250).

Burges, C. J., Shaked, T., Renshaw, E., Lazier, A., Deeds, M. Hamilton, N., et al. (2005). Learning to rank using gradient descent. In *ICML 2005* (pp. 89–96).

Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., & Li, H. (2007). Learning to rank: from pairwise approach to listwise approach. In *ICML 2007* (pp. 129–136).

Ciaramita, M., Murdock, V., & Plachouras, V. (2008). Online learning from click data for sponsored search. In *WWW 2008* (pp. 227–236).

Cossock, D., & Zhang, T. (2006). Subset ranking using regression. In *COLT 2006* (pp. 605–619).

Crammer, K., & Singer, Y. (2002). Pranking with ranking. In *NIPS 2002* (pp. 641–647).

Freund, Y., Iyer, R., Schapire, R. E., & Singer, Y. (2003). An efficient boosting algorithm for combining preferences. *Journal of Machine Learning Research, 4*, 933–969.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *KDD 2002* (pp. 133–142).

Li, P., Burges, C., & Wu, Q. (2007). McRank: Learning to rank using multiple classification and gradient boosting. In *NIPS 2007* (pp. 845–852).

Liu, T.-Y., Xu, J., Qin, T., Xiong, W.-Y., & Li, H. (2007). LETOR: Benchmark dataset for research on learning to rank for information retrieval. In *SIGIR 2007 Workshop on learning to rank for information retrieval*.

Mao, J. (2009). Machine learning in online advertising. In *ICEIS 2009* (pp. IS-27).

Metzler, D. A., & Kanungo, T. (2008). Machine learned sentence selection strategies for query-biased summarization. In *SIGIR 2008 workshop on Learning to Rank for Information Retrieval*.

Surdeanu, M., Ciaramita, M., & Zaragoza, H. (2008). Learning to rank answers on large online qa collections. In *ACL-HLT 2008* (pp. 719–727).

Talyor, M., Guiver, J., et al. (2008). Softrank: Optimising non-smooth rank metrics. In *WSDM 2008* (pp. 77–86).

Verberne, S., Halteren, H. V., Theijssen, D., Raaijmakers, S., & Boves, L. (2009). Learning to rank QA data. In *SIGIR 2009 Workshop on Learning to Rank*.

Xia, F., Liu, T.-Y., Wang, J., Zhang, W., & Li, H. (2008). Listwise approach to learning to rank—theorem and algorithm. In *ICML 2008* (pp. 1192–1199).

Xu, J., Cao, Y., Li, H., & Zhao, M. (2005). Ranking definitions with supervised learning methods. In *WWW 2005* (pp. 811–819). New York, NY, USA: ACM Press.

Yang, Y. H., & Hsu, W. H. (2008a). Video search reranking via online ordinal reranking. In *ICME 2008* (pp. 285–288).

Yang, Y. H., Wu, P. T., Lee, C. W., Lin, K. H., Hsu, W. H., & Chen, H. H. (2008b). Contextseer: context search and recommendation at query time for shared consumer photos. In *ACM Multimedia 2008* (pp. 199–208).

Yue, Y., Finley, T., Radlinski, F., & Joachims, T. (2007). A support vector method for optimizing average precision. In *SIGIR 2007* (pp. 271–278).