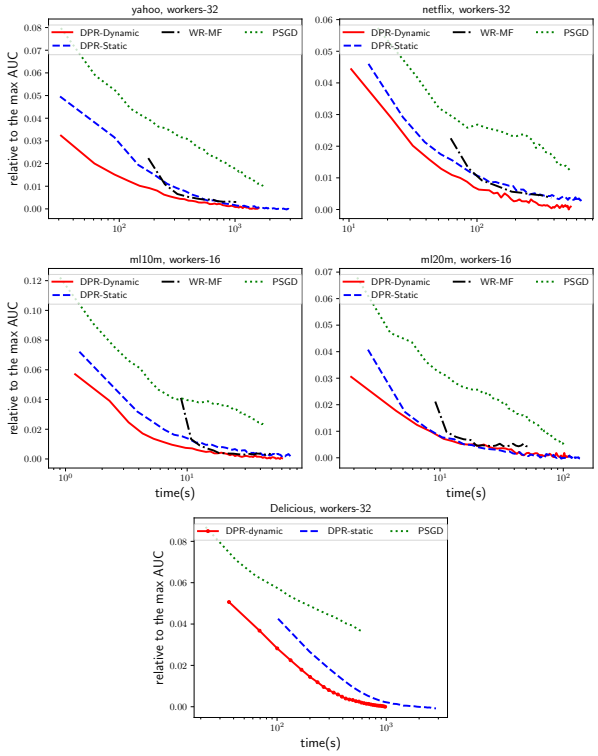

Figure 2: The convergence of different methods on recommender datasets. The y-axis is the relative difference to the maximum AUC value. The x-axis, represents time(sec) in log-scale.

the learning curves recorded by running on a varying number of workers overlap. All the learning curves overlap, which shows the linear scaling of the learning curves across the number of workers.

5 CONCLUSION

In this paper we study the scalability of a pairwise ranking algorithm in distributed settings. We investigate the applicability of distributed SGD techniques for a pointwise method on a pairwise method. The static block partitioning scheme employed by pointwise methods was not useful for pairwise methods. We developed a dynamic block partitioning of model parameters and experimentally show that it works better than the static scheme. The experiments show that DPR outperforms WR-MR, which is a distributed algorithm that optimizes a pointwise loss.

As a future work, it would be interesting to investigate the dynamic partitioning scheme for the exchange of parameters for other model classes as well as in an asynchronous distributed algorithm. This would require overlapping exchange and update steps of the DPR algorithm.

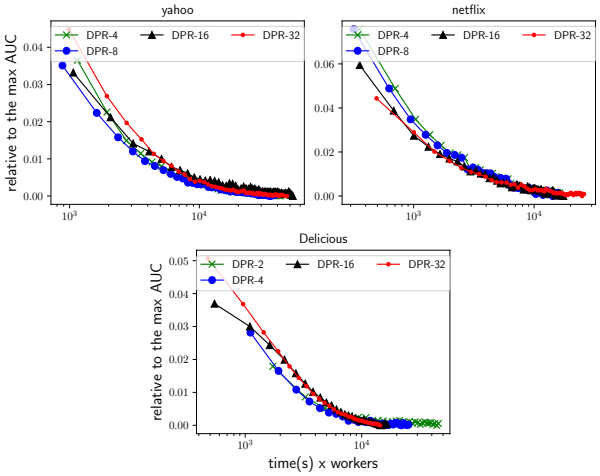

Figure 3: The scalability analysis of DPR (DPR-dynamic) on varying number of workers. The y-axis is the relative difference to the maximum AUC value. The x-axis, represents time(sec) expanded by multiplying by the number of workers in log-scale.

REFERENCES

[1] Robert M. Bell and Yehuda Koren. 2007. Lessons from the Netflix prize challenge. *SIGKDD Explorations* 9, 2 (2007), 75–79.

[2] Gideon Dror, Noam Koenigstein, Yehuda Koren, and Markus Weimer. 2012. The Yahoo! Music Dataset and KDD-Cup '11. In *Proceedings of KDD Cup 2011 competition, San Diego, CA, USA, 2011.* 8–18.

[3] Rainer Gemulla, Erik Nijkamp, Peter J. Haas, and Yannis Sismanis. 2011. Large-scale matrix factorization with distributed stochastic gradient descent. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011.* 69–77.

[4] Olaf Görlitz, Sergej Sizov, and Steffen Staab. 2008. PINTS: peer-to-peer infrastructure for tagging systems. In *Proceedings of the 7th international conference on Peer-to-peer systems, IPTPS'08, Tampa, FL, USA, February 25-26, 2008.* 19.

[5] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative filtering for implicit feedback datasets. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on.* Ieee, 263–272.

[6] Noam Koenigstein and Ulrich Paquet. 2013. Xbox movies recommendations: variational bayes matrix factorization with embedded feature selection. In *Seventh ACM Conference on Recommender Systems, RecSys '13, Hong Kong, China, October 12-16, 2013.* 129–136.

[7] John Levine and Frederick Ducatelle. 2004. Ant colony optimization and local search for bin packing and cutting stock problems. *Journal of the Operational Research Society* 55, 7 (2004), 705–716.

[8] Mu Li, Li Zhou, Zichao Yang, Aaron Li, Fei Xia, David G Andersen, and Alexander Smola. 2013. Parameter server for distributed machine learning. In *Big Learning NIPS Workshop*, Vol. 1.

[9] Steffen Rendle. 2012. Factorization Machines with libFM. *ACM TIST* 3, 3 (2012), 57:1–57:22.

[10] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, UAI '09, Montreal, Quebec, Canada.* 452–461.

[11] A. Murat Yagci, Tevfik Aytekin, and Fikret S. Gürgen. 2017. On Parallelizing SGD for Pairwise Learning to Rank in Collaborative Filtering Recommender Systems. In *Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017.* 37–41.

[12] Lu Yu, Ge Zhou, Chu-Xu Zhang, Junming Huang, Chuang Liu, and Zi-Ke Zhang. 2016. RankMBPR: Rank-Aware Mutual Bayesian Personalized Ranking for Item Recommendation. In *Web-Age Information Management - 17th International Conference, WAIM 2016, Nanchang, China, June 3-5, 2016, Proceedings, Part I.* 244–256.

[13] Erheng Zhong, Yue Shi, Nathan Liu, and Suju Rajan. 2016. Scaling Factorization Machines with Parameter Server. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (CIKM '16).* ACM, New York, NY, USA, 1583–1592. DOI: https://doi.org/10.1145/2983323.2983364

[14] Martin Zinkevich, Markus Weimer, Alexander J. Smola, and Lihong Li. 2010. Parallelized Stochastic Gradient Descent. In *24th Annual Conference on Neural Information Processing Systems, Vancouver, British Columbia, Canada.* 2595–2603.