

Simpson, Edwin; Gurevych, Iryna

Scalable Bayesian preference learning for crowds. (English) Zbl 07224982
Mach. Learn. 109, No. 4, 689-718 (2020).

Summary: We propose a scalable Bayesian preference learning method for jointly predicting the preferences of individuals as well as the consensus of a crowd from pairwise labels. Peoples' opinions often differ greatly, making it difficult to predict their preferences from small amounts of personal data. Individual biases also make it harder to infer the consensus of a crowd when there are few labels per item. We address these challenges by combining matrix factorisation with Gaussian processes, using a Bayesian approach to account for uncertainty arising from noisy and sparse data. Our method exploits input features, such as text embeddings and user metadata, to predict preferences for new items and users that are not in the training set. As previous solutions based on Gaussian processes do not scale to large numbers of users, items or pairwise labels, we propose a stochastic variational inference approach that limits computational and memory costs. Our experiments on a recommendation task show that our method is competitive with previous approaches despite our scalable inference approximation. We demonstrate the method's scalability on a natural language processing task with thousands of users and items, and show improvements over the state of the art on this task. We make our software publicly available for future work (<https://github.com/UKPLab/tacl2018-preference-convincing/tree/crowdGPPL>).

MSC:

68T05 Learning and adaptive systems in artificial intelligence

Software:

BERT; crowdGPPL; GitHub; k-means++; word2vec

Full Text: [DOI](#)

References:

- [1] Abbasnejad, E., Sanner, S., Bonilla, E. V., \& Poupard, P., et al. (2013). Learning community-based preferences via dirichlet process mixtures of Gaussian processes. In *Twenty-third international joint conference on artificial intelligence* (pp. 1213-1219). Retrieved January 17, 2020 from <https://www.ijcai.org/Proceedings/13/Papers/183.pdf>.
- [2] Adams, R. P., Dahl, G. E., \& Murray, I. (2010). Incorporating side information in probabilistic matrix factorization with Gaussian processes. In *Proceedings of the twenty-sixth conference on uncertainty in artificial intelligence* (pp. 1-9). AUAI Press.
- [3] Ahn, S., Korattikara, A., Liu, N., Rajan, S., \& Welling, M. (2015). Large-scale distributed Bayesian matrix factorization using stochastic gradient MCMC. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 9-18). ACM.
- [4] Arthur, D., \& Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on discrete algorithms* (pp. 1027-1035). Society for Industrial and Applied Mathematics. · [Zbl 1302.68273](#)
- [5] Banerji, M.; Lahav, O.; Lintott, CJ; Abdalla, FB; Schawinski, K.; Bamford, SP; Andreescu, D.; Murray, P.; Raddick, MJ; Slosar, A., Galaxy zoo: Reproducing galaxy morphologies via machine learning, *Monthly Notices of the Royal Astronomical Society*, 406, 1, 342-353 (2010)
- [6] Bonilla, E., Steinberg, D., \& Reid, A. (2016). Extended and unscented kitchen sinks. In M. F. Balcan \& K. Q. Weinberger (Eds.), *Proceedings of the 33rd international conference on machine learning*, PMLR, New York, New York, USA, proceedings of machine learning research (Vol. 48, pp. 1651-1659). Retrieved January 17, 2020 from <http://proceedings.mlr.press/v48/bonilla16.html>.
- [7] Bors, AG; Pitas, I., Median radial basis function neural network, *IEEE Transactions on Neural Networks*, 7, 6, 1351-1364 (1996)
- [8] Bradley, RA; Terry, ME, Rank analysis of incomplete block designs: I. The method of paired comparisons, *Biometrika*, 39, 3-4, 324-345 (1952) · [Zbl 0047.12903](#)
- [9] Chen, X., Bennett, P. N., Collins-Thompson, K., \& Horvitz, E. (2013). Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of the sixth ACM international conference on web search and data mining* (pp. 193-202). ACM.
- [10] Chen, G.; Zhu, F.; Heng, PA, Large-scale Bayesian probabilistic matrix factorization with memo-free distributed variational inference, *ACM Transactions on Knowledge Discovery from Data*, 12, 3, 31:1-31:24 (2018)

- [11] Chu, W., \& Ghahramani, Z. (2005). Preference learning with Gaussian processes. In Proceedings of the 22nd international conference on machine learning (pp. 137-144). ACM.
- [12] Devlin, J., Chang, M. W., Lee, K., \& Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Human language technologies (long and short papers), association for computational linguistics, Minneapolis, Minnesota (Vol. 1, pp. 4171-4186). 10.18653/v1/N19-1423.
- [13] Felt, P.; Ringger, E.; Seppi, K., Semantic annotation aggregation with conditional crowdsourcing models and word embeddings, Proceedings of COLING 2016, the 26th international conference on computational linguistics: Technical papers, 1787-1796 (2016), Osaka, Japan: The COLING 2016 Organizing Committee, Osaka, Japan
- [14] Fu, Y.; Hospedales, TM; Xiang, T.; Xiong, J.; Gong, S.; Wang, Y.; Yao, Y., Robust subjective visual property prediction from crowdsourced pairwise labels, IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 3, 563-577 (2016)
- [15] Fürnkranz, J., \& Hüllermeier, E. (2010). Preference learning and ranking by pairwise comparison. In Preference learning (pp. 65-82). Springer. · Zbl 1214.68286
- [16] Gretton, A., Sejdinovic, D., Strathmann, H., Balakrishnan, S., Pontil, M., Fukumizu, K., \& Sriperumbudur, B. K. (2012). Optimal kernel choice for large-scale two-sample tests. In Advances in neural information processing systems (pp. 1205-1213). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/4727-optimal-kernel-choice-for-large-scale-two-sample-tests>.
- [17] Guo, S., Sanner, S., \& Bonilla, E. V. (2010). Gaussian process preference elicitation. In Advances in neural information processing systems (pp. 262-270). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/4141-gaussian-process-preference-elicitation>.
- [18] Habernal, I., \& Gurevych, I. (2016). Which argument is more convincing? analyzing and predicting convincingness of web arguments using bidirectional LSTM. In Proceedings of the 54th annual meeting of the association for computational linguistics (Vol. 1: long papers, pp. 1589-1599). Berlin, Germany: Association for Computational Linguistics.
- [19] Han, B.; Pan, Y.; Tsang, IW, Robust Plackett-Luce model for k-ary crowdsourced preferences, Machine Learning, 107, 4, 675-702 (2018) · Zbl 06889042
- [20] Haykin, S. (2001). Kalman filtering and neural networks. Wiley.
- [21] Hensman, J., Fusi, N., \& Lawrence, N. D. (2013). Gaussian processes for big data. In Proceedings of the twenty-ninth conference on uncertainty in artificial intelligence (pp. 282-290). AUAI Press.
- [22] Hensman, J., Matthews, A. G. D. G., \& Ghahramani, Z. (2015). Scalable variational Gaussian process classification. In Proceedings of the 18th international conference on artificial intelligence and statistics (pp. 351-360). Retrieved January 17, 2020 from <http://proceedings.mlr.press/v38/hensman15>.
- [23] Hoffman, MD; Blei, DM; Wang, C.; Paisley, JW, Stochastic variational inference, Journal of Machine Learning Research, 14, 1, 1303-1347 (2013) · Zbl 1317.68163
- [24] Houlisby, N., Huszar, F., Ghahramani, Z., \& Hernández-Lobato, J. M. (2012). Collaborative Gaussian processes for preference learning. In Advances in neural information processing systems (pp. 2096-2104). Retrieved January 17, 2020 from <http://papers.nips.cc/paper/4700-collaborative-gaussian-processes-for-preference-learning>.
- [25] Joachims, T. (2002). Optimizing search engines using clickthrough data. In Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 133-142). ACM.
- [26] Kamishima, T. (2003). Nantonac collaborative filtering: Recommendation based on order responses. In Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining (pp. 583-588). ACM.
- [27] Kendall, MG, Rank correlation methods (1948), New York City: Griffin, New York City
- [28] Khan, M. E., Ko, Y. J., \& Seeger, M. (2014). Scalable collaborative Bayesian preference learning. In: S. Kaski \& J. Corander (Eds.), Proceedings of the seventeenth international conference on artificial intelligence and statistics, PMLR, Reykjavik, Iceland, proceedings of machine learning research (Vol. 33, pp. 475-483). Retrieved January 17, 2020 from <http://proceedings.mlr.press/v33/khan14>.
- [29] Kim, Y., Kim, W., \& Shim, K. (2014). Latent ranking analysis using pairwise comparisons. In 2014 IEEE international conference on data mining (ICDM), IEEE (pp. 869-874). Retrieved January 17, 2020 from <https://ieeexplore.ieee.org/abstract/document/7023415>.
- [30] Kingsley, DC; Brown, TC, Preference uncertainty, preference refinement and paired comparison experiments, Land Economics, 86, 3, 530-544 (2010)
- [31] Kiritchenko, S., \& Mohammad, S. (2017). Best-worst scaling more reliable than rating scales: A case study on sentiment intensity annotation. In Proceedings of the 55th annual meeting of the association for computational linguistics (Vol. 2: short papers, pp. 465-470). Vancouver, Canada: Association for Computational Linguistics. 10.18653/v1/P17-2074.
- [32] Koren, Y.; Bell, R.; Volinsky, C., Matrix factorization techniques for recommender systems, Computer, 42, 8, 30-37 (2009)
- [33] Lawrence, N. D., \& Urtasun, R. (2009). Non-linear matrix factorization with Gaussian processes. In Proceedings of the 26th international conference on machine learning (pp. 601-608). ACM.
- [34] Lee, SM; Roberts, SJ, Sequential dynamic classification using latent variable models, The Computer Journal, 53, 9, 1415-1429 (2010)
- [35] Li, J., Mantiuk, R., Wang, J., Ling, S., \& Le Callet, P. (2018). Hybrid-MST: A hybrid active sampling strategy for pairwise preference aggregation. In Advances in neural information processing systems (pp. 3475-3485). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/7607-hybrid-mst-a-hybrid-active-sampling-strategy-for-pairwise-preference-aggregation>.
- [36] Lowne, D.; Roberts, SJ; Garnett, R., Sequential non-stationary dynamic classification with sparse feedback, Pattern Recognition, 43, 3, 897-905 (2010) · Zbl 1187.68477
- [37] Luce, RD, On the possible psychophysical laws, Psychological Review, 66, 2, 81 (1959)

- [38] Lukin, S., Anand, P., Walker, M., & Whittaker, S. (2017). Argument strength is in the eye of the beholder: Audience effects in persuasion. In Proceedings of the 15th conference of the European chapter of the association for computational linguistics (pp. 742-753).
- [39] MacKay, DJ, Probable networks and plausible prediction—a review of practical Bayesian methods for supervised neural networks, *Network: Computation in Neural Systems*, 6, 3, 469-505 (1995) · [Zbl 0834.68098](#)
- [40] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/5021-distributed-representations-of-words-and-phrases-and-their-compositionality>.
- [41] Minka, T. P. (2001). Expectation propagation for approximate Bayesian inference. In Proceedings of the seventeenth conference on uncertainty in artificial intelligence (pp. 362-369). arXiv:1301.2294.
- [42] Mo, K., Zhong, E., & Yang, Q. (2013). Cross-task crowdsourcing. In Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 677-685). ACM.
- [43] Mosteller, F., Remarks on the method of paired comparisons: I. The least squares solution assuming equal standard deviations and equal correlations, *Psychometrika*, 16, 3-9 (1951)
- [44] Naish-Guzman, A., & Holden, S. (2008). The generalized FITC approximation. In J. C. Platt, D. Koller, Y. Singer, & S. T. Roweis (Eds.), Advances in neural information processing systems 20 (pp. 1057-1064). Curran Associates, Inc.
- [45] Nguyen, T. V., & Bonilla, E. V. (2014). Collaborative multi-output Gaussian processes. In Proceedings of the thirtieth conference on uncertainty in artificial intelligence (pp. 643-652). AUAI Press.
- [46] Nickisch, H.; Rasmussen, CE, Approximations for binary Gaussian process classification, *Journal of Machine Learning Research*, 9, 2035-2078 (2008) · [Zbl 1225.62087](#)
- [47] Ovadia, S., Ratings and rankings: Reconsidering the structure of values and their measurement, *International Journal of Social Research Methodology*, 7, 5, 403-414 (2004)
- [48] Pan, Y.; Han, B.; Tsang, IW, Stagewise learning for noisy k-ary preferences, *Machine Learning*, 107, 8-10, 1333-1361 (2018) · [Zbl 06990185](#)
- [49] Plackett, RL, The analysis of permutations, *Applied Statistics*, 24, 193-202 (1975)
- [50] Porteous, I., Asuncion, A., & Welling, M. (2010). Bayesian matrix factorization with side information and Dirichlet process mixtures. In Proceedings of the twenty-fourth AAAI conference on artificial intelligence (pp. 563-568). AAAI Press.
- [51] Psorakis, I.; Roberts, S.; Ebden, M.; Sheldon, B., Overlapping community detection using Bayesian non-negative matrix factorization, *Physical Review E*, 83, 6, 066114 (2011)
- [52] Rasmussen, CE; Williams, CKI, *Gaussian processes for machine learning*, 715-719 (2006), Cambridge: The MIT Press, Cambridge
- [53] Reece, S., Roberts, S., Nicholson, D., & Lloyd, C. (2011). Determining intent using hard/soft data and Gaussian process classifiers. In Proceedings of the 14th international conference on information fusion (pp. 1-8). IEEE.
- [54] Resnick, P.; Varian, HR, Recommender systems, *Communications of the ACM*, 40, 3, 56-58 (1997)
- [55] Saha, A., Misra, R., & Ravindran, B. (2015). Scalable Bayesian matrix factorization. In Proceedings of the 6th international conference on mining ubiquitous and social environments (Vol. 1521, pp. 43-54). Retrieved January 17, 2020 from <http://ceur-ws.org/Vol-1521/paper6.pdf>.
- [56] Salakhutdinov, R., & Mnih, A. (2008). Bayesian probabilistic matrix factorization using Markov chain Monte Carlo. In Proceedings of the 25th international conference on machine learning (pp. 880-887). ACM.
- [57] Salimans, T., Paquet, U., & Graepel, T. (2012). Collaborative learning of preference rankings. In Proceedings of the sixth ACM conference on recommender systems (pp. 261-264). ACM.
- [58] Simpson, E., Reece, S., & Roberts, S. J. (2017). Bayesian heatmaps: Probabilistic classification with multiple unreliable information sources. In Joint European conference on machine learning and knowledge discovery in databases (pp. 109-125). Springer.
- [59] Simpson, E.; Gurevych, I., Finding convincing arguments using scalable Bayesian preference learning, *Transactions of the Association for Computational Linguistics*, 6, 357-371 (2018)
- [60] Snelson, E., & Ghahramani, Z. (2006). Sparse Gaussian processes using pseudo-inputs. In Advances in neural information processing systems (pp. 1257-1264). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/2857-sparse-gaussian-processes-using-pseudo-inputs>.
- [61] Snow, R.; O'Connor, B.; Jurafsky, D.; Ng, A., Cheap and fast—but is it good? Evaluating non-expert annotations for natural language tasks, Proceedings of the 2008 conference on empirical methods in natural language processing, 254-263 (2016), Honolulu, Hawaii: Association for Computational Linguistics, Honolulu, Hawaii
- [62] Steinberg, D. M., & Bonilla, E. V. (2014). Extended and unscented Gaussian processes. In Advances in neural information processing systems (pp. 1251-1259). Retrieved January 17, 2020 from <https://papers.nips.cc/paper/5455-extended-and-unscented-gaussian-processes>.
- [63] Thurstone, LL, A law of comparative judgment, *Psychological Review*, 34, 4, 273 (1927)
- [64] Uchida, S., Yamamoto, T., Kato, M. P., Ohshima, H., & Tanaka, K. (2017). Entity ranking by learning and inferring pairwise preferences from user reviews. In Asia information retrieval symposium (pp. 141-153). Springer.
- [65] Vander Aa, T.; Chakroun, I.; Haber, T., Distributed Bayesian probabilistic matrix factorization, *Procedia Computer Science*, 108, 1030-1039 (2017)
- [66] Volkovs, M., Yu, G., & Poutanen, T. (2017). Dropoutnet: Addressing cold start in recommender systems. In I. Guyon, U.

V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, \& R. Garnett (Eds.), *Advances in neural information processing systems* 30 (pp. 4957-4966). Curran Associates, Inc.

- [67] Wang, X., Wang, J., Jie, L., Zhai, C., \& Chang, Y. (2016). Blind men and the elephant: Thurstonian pairwise preference for ranking in crowdsourcing. In *2016 IEEE 16th international conference on data mining (ICDM)* (pp 509-518). IEEE.
- [68] Yang, YH; Chen, HH, Ranking-based emotion recognition for music organization and retrieval, *IEEE Transactions on Audio, Speech, and Language Processing*, 19, 4, 762-774 (2011)
- [69] Yannakakis, G. N., \& Hallam, J. (2011). Ranking vs. preference: A comparative study of self-reporting. In *International conference on affective computing and intelligent interaction* (pp. 437-446). Springer.
- [70] Yi, J., Jin, R., Jain, S., \& Jain, A. (2013). Inferring users' preferences from crowdsourced pairwise comparisons: A matrix completion approach. In *First AAAI conference on human computation and crowdsourcing*. Retrieved January 17, 2020 from <https://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7536>.

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.