指出,用户在对商品打分时关注商品特性多于关注自身的偏好.实验结果表明,HFT(item)效果优于HFT(user).

- TopicMF 方法是更换了主题发现的评论集主体,采用单条评论作为主题模型的输入.而 HFT(user)和HFT(item)的对主题偏好的学习是通过整体评论集使用主题发现,定义了主题偏好分布与用户潜在因子向量和商品潜在因子向量的映射.实验结果表明,TopicMF 方法均优于 HFT(item)和 HFT(user),单条评论能够更精准地表达某用户对某商品的个性偏好.
- 本文提出的 PreferenceMF 是在 TopicMF 基础上加入了主题偏好的引导.经过实验验证,该引导项可以进一步提升推荐质量.后续有待进一步研究该方法在解决冷启动问题的能力.

4 总结和展望

本文提出了融合评分与评论文本的推荐方法,即加入了偏好引导的 PreferenceMF 方法.从不同角度进行分析后,给出了模型的数学表述形式和参数拟合求解的算法.最后,在 28 组 Amazon 数据子集上与多种相关的推荐方法进行实验对比,结果表明,本文提出的 PreferenceMF 方法的推荐质量相对于目前主流的方法有进一步提升.并分析了各个方法的实验结果以及存在的问题.今后可以对隐因子数量和映射关系做进一步研究.本文在融合方法中假设主题发现的主题数量与潜在因子数相等,并且具有相同的权重,可以考虑不同权重的情况.在解决数据稀疏性、模型冷启动和长尾现象等问题上,都有待进一步展开研究.此外,本文是从最大化似然的角度来使推荐模型适应数据,今后可以采用贝叶斯估计,从最大化后验分布的角度来对模型进行建模和求解.

References:

[1] Goldberg D, Nichols D, Oki B M, Terry D. Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992,35(12):61−70.

[2] Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. GroupLens: An open architecture for collaborative filtering of netnews. In: Proc. of the 1994 ACM Conf. on Computer Supported Cooperative Work. New York: ACM Press, 1994. 175−186.

[3] Su C. Recommendation algorithms by exploiting rating matrix and review texts [MS. Thesis]. Wuhan: Wuhan University of Technology, 2016.

[4] Lerman K, Blair-Goldensohn S, McDonald RT. Sentiment summarization: Evaluating and learning user preferences. In: Proc. of the 12th Conf. of the European Chapter of the Association for Computational Linguistics (EACL). Stroudsburg: Association for Computational Linguistics, 2009. 514−522.

[5] Sharma A, Cosley D. Do social explanations work? Studying and modeling the effects of social explanations in recommender systems. In: Proc. of the 22nd Int’l World Wide Web Conf. (WWW). New York: ACM Press, 2013. 1133−1144.

[6] Chen L, Chen G, Wang F. Recommender systems based on user reviews: The state of the art. User Modeling and User-adapted Interaction, 2015,25(2):99−154.

[7] Wen JH, Shu S. Improved collaborative filtering recommendation algorithm of similarity measure. Computer Science, 2014,41(5): 68−71 (in Chinese with English abstract).

[8] Huang X, Qin Z, Chen H. A new user similarity measurement based on a local item space in collaborative filtering recommendation. Journal of Computational Information Systems, 2015,11(10):3501−3508.

[9] Hu Y, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. In: Proc. of the 8th IEEE Int’l Conf. on Data Mining (ICDM). Washington: IEEE, 2008. 263−272.

[10] Basilico J, Hofmann T. Unifying collaborative and content-based filtering. In: Proc. of the 21st Int’l Conf. on Machine Learning (ICML). New York: ACM Press, 2004. 65−72.

[11] Ganu G, Elhadad N, Marian A. Beyond the stars: Improving rating predictions using review text content. In: Proc. of the 12th Int’l Workshop on the Web and Databases. New York: ACM Press, 2009. 1−6.

[12] McAuley J, Leskovec J. Hidden factors and hidden topics: Understanding rating dimensions with review text. In: Proc. of the 7th ACM Conf. on Recommender Systems (RecSys). New York: ACM Press, 2013. 165−172.

[13] Bao Y, Fang H, Zhang J. Topicmf: Simultaneously exploiting ratings and reviews for recommendation. In: Proc. of the 28th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI Press, 2014. 2−8.

[14] Gao YF, Yu WZ, Chao PF, Zheng ZL, Zhang R. Analyzing reviews for rating prediction and item recommendation. Journal of East China Normal University (Natural Science), 2015,(3):80−90 (in Chinese with English abstract).

[15] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 2003, 7(1):76−80.

[16] Choi K, Suh Y. A new similarity function for selecting neighbors for each target item in collaborative filtering. Knowledge-based Systems, 2013,37(1):146−153.

[17] Wu X, Cheng B, Chen JL. Collaborative filtering service recommendation based on a novel similarity computation method. IEEE Trans. on Services Computing, 2017,10(3):352−365.

[18] Koren Y, Bell R, Volinsky C. Matrix factorization techniques for recommender systems. IEEE Computer, 2009,42(8):30 − 37.

[19] Koren Y. Factorization meets the neighborhood: A multifaceted collaborative filtering model. In: Proc. of the 14th ACM Int'l Conf. on Knowledge Discovery and Data Mining (KDD). New York: ACM Press, 2008. 426−434.

[20] Salakhutdinov R, Mnih A. Probabilistic matrix factorization. In: Proc. of the Advances in Neural Information Processing Systems (NIPS). Red Hook: Curran Associates Inc., 2007. 1257−1264.

[21] Xiao R, Li Y, Chen H, *et al*. SRSP-PMF: A novel probabilistic matrix factorization recommendation algorithm using social reliable similarity propagation. In: Proc. of the Intelligent Computing Theories and Methodologies. Springer Int'l Publishing, 2015. 80−91.

[22] Wang C, Blei DM. Collaborative topic modeling for recommending scientific articles. In: Proc. of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining (KDD). New York: ACM Press, 2011. 448−456.

[23] Jordan MI, Blei DM, Ng AY. Latent dirichlet allocation. Journal of Machine Learning Research, 2003,3:993−1022.

[24] DDL, HS.S. Learning the parts of objects by non-negative matrix factorization. Nature, 1999,401(6755):788−791.

[25] Diao Q, Qiu M, Wu CY, Smola AJ, Jiang J. Jointly modeling aspects, ratings and sentiments for movie recommendation (JMARS). In: Proc. of the 20th ACM Sigkdd Int'l Conf. on Knowledge Discovery and Data Mining (KDD). New York: ACM Press, 2014. 193−202.

[26] Ling G, Lyu MR, King I. Ratings meet reviews, a combined approach to recommend. In: Proc. of the 8th ACM Conf. on Recommender Systems (RecSys). New York: ACM Press, 2014. 105−112.

[27] Huang L, Lin CJ, He J, Liu HY, Du XY. Diversified mobile app recommendation combining topic model and collaborative filtering. Ruan Jian Xue Bao/Journal of Software, 2017,28(3):708−720 (in Chinese with English abstract). http://www. jos.org.cn/1000-9825/5163.htm [doi: 10.13328/j.cnki.jos.005163]

[28] Chen X, Yao Y, Xu F, Lu J. Exploring review content for recommendation via latent factor model. In: Proc. of the13th Pacific Rim Int'l Conf. on Artificial Intelligence. Springer Int'l Publishing, 2014. 668−679.

[29] Li H, Lin R, Hong R, Ge Y. Generative models for mining latent aspects and their ratings from short reviews. In: Proc. of the 15th IEEE Int'l Conf. on Data Mining (ICDM). Washington: IEEE, 2015. 241−250.

[30] Song TH, Peng ZH, Wang SZ, Fu WJ, Hong XG, Yu PS. Review-based cross-domain recommendation through joint tensor factorization. In: Proc. of the 22nd Int'l Conf. on Database Systems for Advanced Applications (DASFAA). Springer Int'l Publishing, 2017. 525−540.

[31] Xin X, Liu Z, Lin C, Huang H, Wei X, Guo P. Cross-domain collaborative filtering with review text. In: Proc. of the 24th Int'l Joint Conf. on Artificial Intelligence (IJCAI). Palo Alto: AAAI Press, 2015. 1827−1833.

[32] Zhang HJ, Li ZJ, Chen Y, Zhang XM, Wang SZ. Exploit latent dirichlet allocation for one-class collaborative filtering. In: Proc. of the 23rd ACM Int'l Conf. on Information and Knowledge Management (CIKM). New York: ACM Press, 2014. 1991−1994.

附中文参考文献:

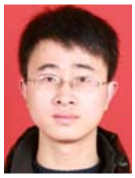
[3] 苏畅.融合评分矩阵和评论文本的推荐算法研究[硕士学位论文].武汉:武汉理工大学,2016.

[7] 文俊浩,舒珊.一种改进相似性度量的协同过滤推荐算法.计算机科学,2014,41(5):68−71.

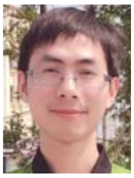
[14] 高祎璠,余文喆,晁平复,郑芷凌,张蓉.基于评论分析的评分预测与推荐.华东师范大学学报(自然科学版),2015,(3):80−90.

[27] 黄璐,林川杰,何军,刘红岩,杜小勇.融合主题模型和协同过滤的多样化移动应用推荐.软件学报,2017,28(3):708−720. http://www.jos.org.cn/1000-9825/5163.htm [doi: 10.13328/j.cnki.jos.005163]

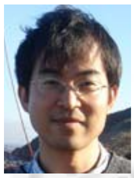
李琳(1977－),女,湖南衡阳人,博士,教授,CCF 专业会员,主要研究领域为数据挖掘,推荐系统,信息检索.



苏畅(1992－),男,硕士,主要研究领域为机器学习,数据挖掘.



朱阁(1993－),男,硕士,主要研究领域为机器学习,数据挖掘.



杨征路(1976－),男,博士,教授,博士生导师,主要研究领域为自然语言处理,人工智能,数据挖掘,信息检索.

解庆(1986－),男,博士,副教授,CCF 专业会员,主要研究领域为数据库,数据挖掘,推荐系统,知识服务.