表示模型、微博用户的静态兴趣及动态兴趣表示及挖掘方法,并给出了微博用户兴趣相似度的计算方法.

不考虑媒体的类型特征,仍然采用统一的用户兴趣表示模型,是不符合实际情况的.而且,挖掘用户的兴趣并不是最终目的,在用户兴趣的基础上,更多的是后期的用户关联度计算、社区划分、信息推荐等应用.本文提出的微博用户的动态兴趣表示模型由于引入了兴趣点的时间分布,可以为兴趣的动态分析提供基础.这种思想同样可以借鉴应用到事件、话题、用户之间的交互等动态特性相关的研究领域.

还需进一步提升的研究内容有:① 微博用户话题提取.由于微博用语过于灵活、不规范,且经常产生新的词语,导致微博话题的提取一直是研究的热点和难点.② 微博网络的海量信息处理.由于微博网络包含了大量的用户及用户生成的微博,有效地分析这类信息需要大数据处理相关技术的支持,包括分布式计算模型、NoSql型的数据存储及检索.③ 不同媒体类型的用户兴趣融合.用户在不同的社交媒体会生成诸多信息,为达到对用户的全面深入分析,需将来源于不同媒体的信息进行融合处理.

References:

[1] Liang YJ, Zheng XL, Zeng DD, Zhou XS, Leischow SJ, Chuang WY. Characterizing social interaction in tobacco-oriented social networks: An empirical analysis. Science Reports, 2015,5(16):1−11. [doi: 10.1038/srep10060]

[2] Wang CX, Guan XH, Qin T, Zhou YD. Modeling on opinion leader's influence in microblog message propagation and its application. Ruan Jian Xue Bao/Journal of Software, 2015,26(6):1473−1485 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4627.htm [doi: 10.13328/j.cnki.jos.004627]

[3] Guo L, Ma J, Chen ZM, Jiang HR. Incorporating item relations for social recommendation. Chinese Journal of Computers, 2014, 37(1):219−228 (in Chinese with English abstract).

[4] Wang XF, Tang L, Gao HJ, Liu H. Discovering overlapping groups in social media. In: Proc. of the 10th IEEE Int'l Conf. on Data Mining. IEEE Computer Society, 2010. 569−578. [doi: 10.1109/ICDM.2010.48]

[5] Diaby M, Viennet E, Launay T. Exploration of methodologies to improve job recommender systems on social networks. Social Network Analysis and Mining, 2014,4(227):1−17. [doi: 10.1007/s13278-014-0227-z]

[6] Ma H, Zhou D, Liu C, Lyu MR, King I. Recommender systems with social regularization. In: Proc. of the 4th ACM Int'l Conf. on Web Search and Data Mining (WSDM 2011). New York: ACM, 2011. 287−296. [doi: 10.1145/1935826.1935877]

[7] Kantor PB, Ricci F, Rokach L, Shapira B. Recommender Systems Handbook. New York: Springer-Verlag, 2009.

[8] Tsai WH, LinYT, Lee KR. Development of social-aware recommendation system using public preference mining and social influence analysis: A case study of landscape recommendation. Journal of Internet Technology, 2016,17(3):561−569. [doi: 10.6138/JIT.2016.17.3.20151110a]

[9] Cruz JD, Bothorel C, Poulet F. Entropy based community detection in augmented social networks. In: Proc. of the Int'l Conf. on Computational Aspects of Social Networks. 2011. 163−168. [doi: 10.1109/CASON.2011.6085937]

[10] Qi GJ, Aggarwal CC, Huang T. Community detection with edge content in social media networks. In: Proc. of the Int'l Conf. on Data Engineering. 2012. 534−545. [doi: 10.1109/ICDE.2012.77]

[11] Ghosh S, Sharma N, Benevenuto F, Ganguly N, Gummadi KP. Cognos: Crowdsourcing search for topic experts in microblogs. In: Proc. of the 35th Int'l ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR). New York, 2012. 575−590. [doi: 10.1145/2348283.2348361]

[12] Liang C, Liu ZY, Sun MS. Expert finding for microblog misinformation identification. In: Proc. of the 24th ACL Int'l Conf. on Computational Linguistics. Mumbai, 2012. 703−712.

[13] Akcora CG, Carminati B, Ferrari E. User similarities on social networks. Social Network Analysis and Mining, 2013,3(3):475−495. [doi: 10.1007/s13278-012-0090-8]

[14] Xing QL, Liu L, Liu YQ, Zhang M, Ma SP. Study on user tags in Weibo. Ruan Jian Xue Bao/Journal of Software, 2015,26(7): 1626−1637 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4655.htm [doi: 10.13328/j.cnki.jos.004655]

[15] Wang X, Jia Y, Zhou B, Chen RH, Han Y. Interaction relation based user tag prediction in microblog site. Computer Engineering & Science, 2013,35(10):44−50 (in Chinese with English abstract).

[16] Ma YF, Zeng Y, Ren X, Zhong N. User interests modeling based on multi-source personal information fusion and semantic reasoning. In: Proc. of the 7th Int'l Conf. on Active Media Technology (AMT 2011). Berlin, Heidelberg: Springer-Verlag, 2011. 195−205. [doi: 10.1007/978-3-642-23620-4_23]

[17] Chen JL, Nairn R, Nelson L, Bernstein M, Chi EH. Short and tweet: Experiments on recommending content from information streams. In: Proc. of the SIGCHI Conf. on Human Factors in Computing Systems (CHI 2010). New York: ACM, 2010. 1185−1194. [doi: 10.1145/1753326.1753503]

[18] Weng JS, Lim EP, Jiang J, He Q. TwitterRank: Finding topic-sensitive influential Twitterers. In: Proc. of the 3rd ACM Int'l Conf. on Web Search and Data Mining. New York, 2010. 261−270. [doi: 10.1145/1718487.1718520]

[19] Zhao WX, in, Jiang J, Weng JS, He J, Lim EP, Yan HF, Li XM. Comparing Twitter and traditional media using topic models. In: Proc. of the 33rd European Conf. on Information Retrieval. Berlin, Heidelberg: Springer-Verlag, 2011. 338−349. [doi: 10.1007/978-3-642-20161-5_34]

[20] Zhou XP, Liang X, Zhang HY. User community detection on micro-blog using R-C model. Ruan Jian Xue Bao/Journal of Software, 2014,25(12):2808−2823 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4720.htm [doi: 10.13328/j.cnki.jos.004720]

[21] Steyvers M, Smyth P, Rosen-Zvi M, Griffiths T. Probabilistic author-topic models for information discovery. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2004. 306−315. [doi: 10.1145/1014052.1014087]

[22] Zhang ZF, Li QD, Zeng D, Gao H. User community discovery from multi-relational networks. Decision Support Systems, 2013, 54(2):870−879. [doi: 10.1016/j.dss.2012.09.012]

[23] Peng ZH, Sun L, Han XP, Shi B. Microblog user recommendation using learning to rank. Journal of Chinese Information Processing, 2013,27(4):96−102 (in Chinese with English abstract).

[24] Hong LJ, Davison BD. Empirical study of topic modeling in Twitter. In: Proc. of the 1st Workshop on Social Media Analytics. Washington, 2010. 80−88. [doi: 10.1145/1964858.1964870]

[25] Hu Y, Wang CJ, Wu J, Xie JY, Li H. Overlapping community discovery and global representation on microblog network. Ruan Jian Xue Bao/Journal of Software, 2014,25(12):2824−2836 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4721.htm [doi: 10.13328/j.cnki.jos.004721]

[26] Zhao X, Chen RS, Fan K, Yan HF, Li XM. A novel burst-based text representation model for scalable event detection. In: Proc. of the 50th Annual Meeting of the Association for Computational Linguistics. 2012. 43−47.

[27] He M, Wang LH, Du P, Zhang J, Cheng XQ. Microblog hot topic detection based on meaningful string clustering. Journal on Communications, 2013,34(Z1):256−262 (in Chinese with English abstract).

[28] He M, Du P, Zhang J, Liu Y, Cheng XQ. Microblog bursty topic detection method based on momentum model. Journal of Computer Research and Development, 2015,52(5):1022−1028 (in Chinese with English abstract).

[29] Shen GW, Yang W, Wang W, Yu M. Burst topic detection oriented large-scale microblog streams. Journal of Computer Research and Development, 2015,52(2):512−521 (in Chinese with English abstract).

[30] Peng ZH, Sun L, Han XP, Chen B. Community hot statuses recommendation. Journal of Computer Research and Development, 2015,52(5):1014−1021 (in Chinese with English abstract).

[31] Xu ZM, Li D, Liu T, Li S, Wang G, Yuan SL. Measuring similarity between microblog users and its application. Chinese Journal of Computers, 2014,37(1):207−218 (in Chinese with English abstract).

[32] Zhang J, Gao JF, Zhou M. Extraction of Chinese compound words: An experimental study on a very large corpus. In: Proc. of the 2nd Workshop on Chinese Language Processing: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics. 2000. 132−139.

[33] Yang WS, Luo AM, Zhang MM. Trust-Circle based recommendation on user cold-start. Computer Science, 2013,40(11a):363−366 (in Chinese with English abstract).

附中文参考文献:

[2] 王晨旭,管晓宏,秦涛,周亚东.微博消息传播中意见领袖影响力建模研究.软件学报,2015,26(6):1473−1485. http://www.jos.org.cn/1000-9825/4627.htm [doi: 10.13328/j.cnki.jos.004627]

[3] 郭磊,马军,陈竹梅,姜浩然.一种结合推荐对象间关联关系的社会化推荐算法.计算机学报,2014,37(1):219−228.

[14] 邢千里,刘列,刘奕群,张敏,马少平.微博中用户标签的研究.软件学报,2015,26(7):1626−1637. http://www.jos.org.cn/1000-9825/4655.htm [doi: 10.13328/j.cnki.jos.004655]

[15] 汪祥,贾焰,周斌,陈儒华,韩毅.基于交互关系的微博用户标签预测.计算机工程与科学,2013,35(10):44−50.

[20] 周小平,梁循,张海燕.基于 R-C 模型的微博用户社区发现.软件学报,2014,25(12):2808−2823. http://www.jos.org.cn/1000-9825/4720.htm [doi: 10.13328/j.cnki.jos.004720]

[23] 彭泽环,孙乐,韩先培,石贝.基于排序学习的微博用户推荐.中文信息学报,2013,27(4):96−102.

[25] 胡云,王崇骏,吴骏,谢俊元,李慧.微博网络上的重叠社群发现与全局表示.软件学报,2014,25(12):2824−2836. http://www.jos.org.cn/1000-9825/4721.htm [doi: 10.13328/j.cnki.jos.004721]

[27] 贺敏,王丽宏,杜攀,张瑾,程学旗.基于有意义串聚类的微博热点话题发现方法.通信学报,2013,34(Z1):256−262.

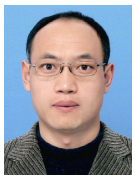
[28] 贺敏,杜攀,张瑾,刘悦,程学旗.基于动量模型的微博突发话题检测方法.计算机研究与发展,2015,52(5):1022−1028.

[29] 申国伟,杨武,王巍,于淼.面向大规模微博消息流的突发话题检测.计算机研究与发展,2015,52(2):512−521.

[30] 彭泽环,孙乐,韩先培,陈波.社区热点微博推荐研究.计算机研究与发展,2015,52(5):1014−1021.

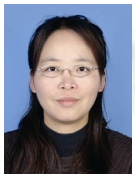
[31] 徐志明,李栋,刘挺,李生,王刚,袁树仑.微博用户的相似性度量及其应用.计算机学报,2014,37(1):207−218.

[33] 杨圩生,罗爱民,张萌萌.基于信任环的用户冷启动推荐.计算机科学,2013,40(11a):363−366.



仲兆满(1977－),男,江苏赣榆人,博士,副教授,主要研究领域为信息检索,文本信息挖掘,事件本体.

胡云(1978－),女,博士,副教授,CCF 专业会员,主要研究领域为数据挖掘,复杂网络.



管燕(1976－),女,讲师,主要研究领域为数据挖掘,模式识别.

李存华(1963－),男,博士,教授,主要研究领域为数据挖掘,人工智能,图像处理.