

大,则说明系统性能越好.

3.2 基线实验

本文选取了以下的方法作为基线实验:

- **MostPopularity(流行度)**.在很多产品推荐系统中,基于流行度的方法是经常使用的基线实验.流行度算法按照物品(类别)的流行度给用户推荐其最热门的几种物品/类别.
- **SVM**.为了证明MLKNN分类器在隐式消费意图识别中的有效性,我们将其与SVM分类器进行了对比.本实验使用的SVM分类器为libsvm-2.82.我们利用本文提出的4类特征在自动标注的数据集上对SVM分类器进行了实验(这里使用了libsvm-2.82默认的RBF核函数).

3.2.1 对隐式消费意图识别结果的评价

由于国内外没有公开发布的隐式消费意图评价测试语料,本文利用自动标注的语料进行评价.具体来说,我们将全部链指的91369个用户数据平均分为2份,其中一份用于训练,剩余一份用于测试.本文对用户的隐式消费意图类别进行了统计分析,发现90%以上的隐式消费意图类别在1类以上,用户的平均隐式消费意图类别在4类.这说明识别出用户的隐式消费意图类别是分析用户隐式消费意图的一个重要的基本单元.因此,隐式消费意图类别的正确识别有助于消费意图相关任务的解决.

本节共对3种方法进行了实验对比,包括:(a) 基准系统;(b) SVM分类方法;(c) MLKNN分类方法.其中,基准系统是指根据购买类别的流行度进行排序,即对在训练数据集中的用户购买类别的情况进行排序.

表5给出了3种方法在测试数据集上的评测结果.其中,基准系统只给出了51.2%的平均精度,这说明隐式消费意图识别并不是一个简单的任务.为了证明MLKNN分类器在隐式消费意图识别中的有效性,我们将其与SVM分类器进行了对比.从表5中可以看到,两类分类器的性能都远高于基准系统,其中,SVM给出了72.3%的平均精度,这表明利用提出的全部特征自动在训练数据集上学习得到的分类器是有效的.利用MLKNN(取 $k=10$)的方法学习,使其得到了更好的性能.

Table 5 Experimental results of different approaches for implicit consumption intent

表 5 各种隐式消费意图识别方法的对比实验结果

评价指标		Baseline	SVM	MLKNN
基于实例	Hanmming loss↓	0.383	0.264	0.147
	One-Error↓	0.094	0.026	0.011
基于排序	Coverage↓	8.982	7.625	6.656
	Ranking loss↓	0.179	0.126	0.085
	AvePrecision↑	0.512	0.723	0.835

3.2.2 对分类特征的评价

首先,利用上述标注数据对本文提出的分类特征进行评价.为考察本文使用的4类特征是否对隐式消费意图识别都有作用,我们进行了4组实验,每组实验依次加入基于用户关注行为特征(特征1)、用户意图关注行为特征(特征2)、用户意图转发行为特征(特征3)以及用户性别特征(特征4).其实验结果见表6.从表6中我们可以看到,随着每一类特征的加入,分类的平均精度都有明显的提高,其他4项评价指标都有明显的降低.尤其是当使用全部4类特征时,分类的平均精度达到最高,其他4项指标达到最低.这说明本文所采用的4类特征对于提高多元分类的性能都是有帮助的.也就是说,全部4类特征均有助于隐式消费意图的识别.

Table 6 Contributions of the 4 kinds of features

表 6 4类特征的贡献

特征贡献	Hanmming loss	One-Error	Coverage	Ranking loss	AvePrecision
特征1	0.318	0.028	7.835	0.240	0.693
特征1+2	0.194	0.017	6.963	0.172	0.817
特征1+2+3	0.183	0.015	6.937	0.106	0.821
特征1+2+3+4	0.147	0.011	6.656	0.085	0.835

4 结论与展望

本文提出了一种社交媒体用户中的隐式消费意图识别方法,并将隐式消费意图识别作为一个多标签分类问题加以解决.具体来说,本文利用 MLKNN 分类器解决隐式消费意图中的多分类问题,并综合使用了 4 类特征,即:(1) 用户关注行为特征;(2) 用户个人信息特征;(3) 用户消费意图转发行为特征;(4) 用户消费意图关注行为特征.在隐式消费意图识别评价方面,本文尝试将不同媒体用户中的用户身份链接起来,自动抽取了 12 万余对社交媒体网站和电子商务网站的用户数据.在此自动评价集上的实验结果表明,本文使用的分类器和 4 类特征对于隐式消费意图识别都是有效的.

在今后的工作中,我们会尝试寻找一种方法来标定用户隐式消费意图中真正希望购买的具体产品,将更加方便对用户迅速推荐所需的产品信息.此外,我们也会考虑将本文提出的自动链指方法应用到其他不同社区中,将一个用户的不同形式、全面的信息都聚合起来,形成更丰富的个人信息,从而解决推荐系统和个性化系统中的冷启动问题.

致谢 在此,我们向对本研究工作提供帮助的老师和同学表示感谢.特别要感谢李一鸣、焦阳等同学在实验数据处理上的工作,还要感谢宋巍、伍大勇、张伟男等同学对本文初稿进行审阅并提出宝贵意见.

References:

- [1] Goldberg AB, Fillmore N, Andrzejewski D, Xu Z, Gibson B, Zhu XJ. May all your wishes come true a study of wishes how to recognize them. In: Proc. of the 2009 Annual Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (HLT-NAACL). Morristown: ACL Press, 2009. 263–271.
- [2] Chen ZY, Liu B, Hsu M, Castellanos M, Ghosh R. Identifying intention posts in discussion forums. In: Proc. of the 2013 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT), Morristown: ACL Press, 2013. 1041–1050.
- [3] Fu B, LIU T. Weakly-Supervised consumption intent detection in microblogs. Journal of Computational Information Systems, 2013, 6(9):2423–2431.
- [4] Zhang Y, Pennacchiotti M. Predicting purchase behaviors from social media. In: Proc. of the Int'l World Wide Web Conf. Steering Committee (WWW). New York: ACM Press, 2013. 1521–1532. [doi: 10.1145/2488388.2488521]
- [5] Ding X, Liu T, Duan JW, Nie JY. Mining user consumption intention from social media using domain adaptive convolutional neural network. In: Proc. of the Association for the Advancement of Artificial Intelligence (AAAI). Menlo Park: AAAI Press, 2015. 2389–2395.
- [6] Liu J, Zhang F, Song XY. What's in a name? An unsupervised approach to link users across communities. In: Proc. of the 6th ACM Int'l Conf. on Web Search and Data Mining (WSDM). New York: ACM Press, 2013. 495–504. [doi: 10.1145/2433396.2433457]
- [7] Chiu YB, Lin CP, Tang LL. Gender differs: Assessing a model of online purchase intentions in e-tail service. Int'l Journal of Service Industry Management, 2005,16(5):416–435. [doi: 10.1108/09564230510625741]
- [8] Zhang ML, Zhou ZH. A k -nearest neighbor based algorithm for multi-label classification. In: Proc. of the 1st IEEE Int'l Conf. on Granular Computing (GrC). IEEE Press, 2005. 718–721. [doi: 10.1109/GRC.2005.1547385]
- [9] Tsoumakas G, Katakis I. Multi-Label classification: An overview. Int'l Journal of Data Warehousing and Mining, 2007,3(3):1–13. [doi: 10.4018/jdwm.2007070101]
- [10] Jiang Y, She QQ, Li M, Zhou ZH. A transductive multi label text categorization approach. Journal of Computer Research and Development, 2008,45(11):1817–1823 (in Chinese with English abstract).
- [11] Schapire RE, Singer Y. Boostexter: A boosting-based system for text categorization. Machine Learning, 2000,39(2):135–168. [doi: 10.1023/A:1007649029923]
- [12] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine Learning, 2011,85(3):333–359. [doi: 10.1007/s10994-011-5256-5]

- [13] Tsoumakas G, Katakis I, Vlahavas I. Mining multi-label data. In: Data Mining and Knowledge Discovery Handbook. 667–685. [doi: 10.1007/978-0-387-09823-4_34]
- [14] He ZF, Yang M, Liu HD. Joint learning of multi-label classification and label correlations. Ruan Jian Xue Bao/Journal of Software, 2014,25(9):1967–1981 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4634.htm> [doi: 10.13328/j.cnki.jos.004634]
- [15] Chen Y, Lin L, Sun CJ, Liu BQ. A tag recommendation method for microblog users. Intelligent Computer and Applications, 2011, 1(5):21–26 (in Chinese with English abstract). [doi: 10.3969/j.issn.2095-2163.2011.05.006]
- [16] Jin YA, Li RX, Wen KM, Gu XW, Lu ZD, Duan DS. A survey on social annotation and its application in information retrieval. Journal of Chinese Information Processing, 2010,24(4):52–62 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0077.2010.04.008]
- [17] Zhang Y, Pennacchiotti M. Recommending branded products from social media. In: Proc. of the 7th ACM Conf. on Recommender Systems. New York: ACM Press, 2013. 77–84. [doi: 10.1145/2507157.2507170]

附中文参考文献:

- [10] 姜远,余俏俏,黎铭,周志华.一种直推式多标记文档分类方法.计算机研究与发展,2008,45(11):1817–1823.
- [14] 何志芬,杨明,刘会东.多标记分类和标记相关性的联合学习.软件学报,2014,25(9):1967–1981. <http://www.jos.org.cn/1000-9825/4634.htm> [doi: 10.13328/j.cnki.jos.004634]
- [15] 陈渊,林磊,孙承杰,刘秉权.一种面向微博用户的标签推荐方法.智能计算机与应用,2011,1(5):21–26. [doi: 10.3969/j.issn.2095-2163.2011.05.006]
- [16] 靳延安,李瑞轩,文坤梅,辜希武,卢正鼎,段东圣.社会标注及其在信息检索中的应用研究综述.中文信息学报,2010,24(4):52–62. [doi: 10.3969/j.issn.1003-0077.2010.04.008]



付博(1983—),女,黑龙江海伦人,博士,主要研究领域为社会计算,信息检索.



刘挺(1972—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为社会计算,信息检索,自然语言处理.