



基于潜在因子模型的跨领域信息推荐算法

高升,任思婷,郭军

(北京邮电大学 北京 100876)

摘要:互联网环境下,不同领域中多源异构信息对象的交互融合使用户面临大数据环境下的信息选择困境,传统的信息推荐算法已很难适应跨领域的信息推荐服务。综合分析了不同领域内用户对信息对象的评价数据,基于潜在因子模型抽取了不同领域中某一用户聚类集合对某一信息对象聚类集合评分模式的跨领域共性特征和单领域个性特征,进而通过传递、共享跨域共性特征信息的方式缓解了目标领域的数据稀疏性问题,提高了跨域信息推荐的准确度。

关键词:跨领域信息推荐;潜在因子模型;用户评价模式

doi: 10.11959/j.issn.1000-0801.2015188

Cross-Domain Recommendation Algorithm Based on Latent Factor Model

Gao Sheng, Ren Siting, Guo Jun

(Beijing University of Posts and Telecommunications, Beijing 100876, China)

Abstract: In the internet environment, the combining of multi-source heterogeneous information objects in different areas makes users face information selection dilemma problem in big data environment. It has been very difficulty for traditional information recommendation algorithms to adapt to the interdisciplinary information recommendation service. The evaluation model from a user clustering set to an information object clustering set has common characteristics of cross-domain and personality characteristics of single domain. By analyzing the evaluation data from users to information objects in different areas, these characteristics were extracted based on latent factor model. Then by transmitting and sharing the common characteristics of cross-domain, the data sparseness problem of target field was alleviated, which could improve the accuracy of cross-domain information recommendation.

Key words: cross-domain recommendation, latent factor model, user rating pattern

1 引言

随着互联网和 Web2.0 技术的发展,社交网络、微博、

维基、视频共享等各类在线社交应用迅速增长,为用户的信息传播与共享、个人意见表达、情感交流和经济互动创建了丰富的通信平台,特别是随着互联网信息资源规模的

收稿日期:2015-05-10;修回日期:2015-07-15

基金项目:国家自然科学基金资助项目(No.61300080, No.61273217),教育部—中国移动科研基金资助项目(No. MCM20130310)

Foundation Items: The National Science Foundation of China (No.61300080, No.61273217), Ministry of Education-China Mobile Research Foundation (No.MCM20130310)

论文引用格式:高升,任思婷,郭军. 基于潜在因子模型的跨领域信息推荐算法. 电信科学, 2015188

Gao S, Ren S T, Guo J. Cross-domain recommendation algorithm based on latent factor model. Telecommunications Science, 2015188



迅猛增长,用户需要在海量的文本、图像、音/视频等大数据中获取符合个性化需求的信息也愈加困难,从而导致了“信息过载”问题。信息推荐系统作为一种智能信息处理技术,可借助用户历史行为数据以及用户评价数据对用户各种信息需求、兴趣及偏好进行描述,并依据用户兴趣模型和信息对象描述模型通过一定的智能推荐算法主动实现信息对象的选择,最终将符合用户兴趣的信息对象以某种展现形式提供给目标用户。

然而,互联网环境下,不同领域中多源异构信息对象的交互融合使用户面临大数据环境下的信息选择困境,传统的信息推荐系统及其核心算法已很难适应用户规模、信息推荐项目数量和用户历史评价数据的快速增长,仅仅针对单一特定目标领域内的信息对象提供推荐服务^[1],例如,Netflix 系统只向用户推荐电影和电视信息;YouTube 网站只提供视频领域的信息推荐等。传统推荐系统无法灵活扩展到对不同领域中对信息对象进行推荐,例如,推荐系统无法向不同行业领域中的非重合用户群体同时推荐相关的电影和科技文献。事实上,互联网环境下,每个行业领域都积累了大量的商业及信息评价等数据,但是信息对象内容的多样性、用户群体的偏好性以及不同领域知识的相对独立性都使得网络资源的有效利用率很低,无法实现跨领域信息资源的共享和充分利用;同时,用户对海量信息对象的评价数据稀疏性问题、新用户或新商品的“冷启动”问题、推荐算法的扩展性问题都造成推荐质量明显下降^[1],严重制约了传统推荐系统的进一步推广和应用。因此,本文提出了跨领域信息推荐这一新问题,该问题源于实际信息需求,主要针对跨领域中的非重合用户群体,不仅可以向其推荐符合其个性化需求的所属领域的信息对象,还可以推荐跨领域的不同类型的信息对象,从而提高跨领域信息对象推荐结果的准确性和多样性,提升跨领域信息资源的利用率。特别地,通过综合分析不同领域内用户对信息对象的评价数据,基于潜在因子模型抽取了不同领域中某一用户聚类集合对某一信息对象聚类集合评分模式的跨领域共性特征和单领域个性特征,进而通过传递、共享跨领域共性特征信息的方式来缓解目标领域的数据稀疏性问题,以提高跨域信息推荐的准确度。

2 相关工作

信息推荐算法作为推荐系统的核心技术已经成为个性化信息服务的研究重点。对于单领域信息推荐系统来

说,目前效果最好的信息推荐算法是基于矩阵分解的潜在因子模型(latent factor model,LFML)^[2]算法。潜在因子模型是一种基于机器学习的方法,具有较好的理论基础:通过构建样本数据的低维潜在空间,将高维原始数据映射到潜在空间中,获取样本点的潜在因子,从而捕捉数据的内隐特征和本质几何结构。在信息推荐系统的应用中,潜在因子模型一般用潜在因子表达用户和信息对象的潜在特征,其作为一种实现大数据处理与分析的有效工具,已在很多领域取得了广泛的应用,如应用于电影信息推荐领域的SVD++(singular value decomposition++)模型^[2]、应用于数据挖掘领域的LDA(latent dirichlet allocation)模型^[3]以及应用于信息检索领域的PLSI(probabilistic latent semantic indexing)模型^[4,5]等。这些潜在因子模型在学习样本数据的潜在因子时多采用了经典矩阵分解算法,如主成分分析(principle component analysis,PCA)^[6]算法、奇异值分解(singular value decomposition,SVD)^[6]算法、非负矩阵分解(nonnegative matrix factorization,NMF)^[7]算法以及稀疏编码(sparse coding,SC)^[8]算法等。然后,模型通过在潜在空间中构建的潜在特征因子内积值来预测目标用户对潜在推荐对象的喜好程度,从而实现信息推荐。然而,潜在因子模型一般用于处理单领域内容的信息特征挖掘,并不适用于处理跨领域多源异构信息建模及信息推荐。

跨领域信息推荐问题在国内外开展的研究还比较少^[9]。国内清华大学、复旦大学、上海交通大学以及香港科技大学的相关研究团队多是从单领域潜在因子模型在重合用户群体信息推荐中的应用角度开展研究,其中,Pan W 等^[10]提出的基于迁移学习的协同过滤模型主要用于对不同领域中相同用户群的行为进行建模,无法解决跨域情况下非重合用户群的兴趣建模和信息推荐问题;Li B 等^[11]提出的模型可以抽取出不同领域中用户群体对信息对象评价模式的共性特征,从而实现跨域信息共享,但是无法分辨出不同领域中用户评价模式的个性特征,而且最终实现的仍是单领域信息推荐;国外研究团队中,Singh 和 Gordan^[12]提出了联合矩阵分解模型来实现关系学习,模型同样针对不同领域中的相同用户集合应用了潜在因子学习,而无法推广到实际跨领域中的不同用户群体和信息对象的情况,应用范围较窄;Berkovsky 和 Shi 等^[13]提出的贝叶斯概率模型可以分析不同领域中评分数据分布的相关性,进而通过传递这种相关性实现跨域信息推荐,但无法实现对用户行为和兴趣的建模,并且算法缺少处理大规模数据

的可扩展性。

3 基于潜在因子模型的跨领域信息推荐算法

跨领域信息推荐问题定义如下:假设给定 Z 个不同的领域,其中第 z 个领域中的一组用户 $U_z=\{u_1^{(z)}, \dots, u_{m_z}^{(z)}\}$ 对一组信息对象 $V_z=\{v_1^{(z)}, \dots, v_{n_z}^{(z)}\}$ 进行评分,产生的评分数据可以表示为一系列的三元组 $D_z=\{(u_1^{(z)}, v_1^{(z)}, r_1^{(z)}), \dots, (u_{m_z}^{(z)}, v_{n_z}^{(z)}, r_{n_z}^{(z)})\}$,其中, S_z 表示第 z 个领域中用户评分的数量,每个领域中的用户评分 r 取值范围都一致,如 1~5 分,但是每个领域中的用户集合或信息对象集合可以有重合或完全不重合。通过综合分析不同领域内用户对信息对象的评价数据,基于潜在因子模型抽取不同领域中某一用户聚类集合对某一信息对象聚类集合评分模式的跨领域共性特征和单领域个性特征,进而通过传递、共享跨域共性特征信息的方式来缓解特定目标领域的信息稀疏性问题,以提高跨域信息推荐的准确度。

3.1 基于用户聚类的潜在因子

信息推荐系统是建立在用户个性化信息需求或兴趣偏好基础上的信息服务系统,关于用户的信息需求、兴趣偏好和使用习惯等知识是推荐算法实现的重要依据,而且用户在选择信息对象时会表现出多种兴趣偏好或者消费购买习惯。因此,在采用基于聚类的潜在因子模型对各个领域内的用户进行聚类时,假定每个领域中的用户都可以从属多个类别,即 $\{C_u^{(1)}, C_u^{(2)}, \dots, C_u^{(K)}\}$ 共 K 个用户集合,每个用户属于一个集合的概率用 $P(C_u^{(k)}|u)$ 表示。构建以用户聚类为中心的潜在因子来表示用户的不同兴趣特征,并体现用户的相关聚类信息,表示为:

$$p_u=[P(C_u^{(1)}|u), P(C_u^{(2)}|u), \dots, P(C_u^{(K)}|u)] \quad \text{s.t. } p_u \mathbf{1}=1 \quad (1)$$

3.2 基于信息对象聚类的潜在因子

类似地,在采用基于聚类的潜在因子模型对各个领域内的信息对象进行聚类时,同样假定每个领域中的信息对象可以从属于多个类别,但是,考虑到不同领域的信息对象可能具有不同的属性特征,为了进一步精确地刻画信息对象的聚类特征,假定每个领域的信息对象的类别信息可以具有跨领域的共性聚类和各领域的特定聚类信息,分别表示如下。

- 针对每个领域的信息对象所具有的跨领域共性聚类,假定 Z 个领域的信息对象都具有 T 个聚类

$\{C_{v-com}^{(1)}, C_{v-com}^{(2)}, \dots, C_{v-com}^{(T)}\}$ 。因此,基于跨领域共性聚类信息的信息对象潜在因子可以表示为:

$$p_{v-com}=[P(C_{v-com}^{(1)}|v), P(C_{v-com}^{(2)}|v), \dots, P(C_{v-com}^{(T)}|v)]$$

$$\text{s.t. } p_{v-com} \mathbf{1}=1 \quad (2)$$

- 针对每个领域的信息对象所具有的特定领域的聚类信息,假定第 z 个领域中的信息对象具有 L_z 个特定领域的聚类 $\{C_{v-spe}^{(1)}, C_{v-spe}^{(2)}, \dots, C_{v-spe}^{(L_z)}\}$ 。因此,基于各领域特定聚类信息的潜在因子可以表示为:
- $$p_{v-spe}^z=[P(C_{v-spe}^{(1)(z)}|v), P(C_{v-spe}^{(2)(z)}|v), \dots, P(C_{v-spe}^{(L_z)}|v)]$$
- $$\text{s.t. } p_{v-spe} \mathbf{1}=1 \quad (3)$$

3.3 基于潜在因子模型的跨领域推荐算法

通过潜在因子模型抽取不同领域中某一用户聚类对某一信息对象聚类的评分模式,并基于信息对象所具有的跨领域共性聚类和各领域特定聚类信息,进一步提取出用户聚类集合的评分模式所体现的跨领域共性特征和单领域特定特征,分别表示为 $P(r|C_u^{(i)}, C_{v-com}^{(j)})$ 和 $P(r|C_u^{(i)}, C_{v-com}^{(j)(z)})$ 。

对于 Z 个领域来说,进一步构建出基于跨领域共性特征的用户评分模式,表示为:

$$S_{com} = \begin{bmatrix} \sum_r r P(r|C_u^{(1)}, C_{v-com}^{(1)}) & \dots & \sum_r r P(r|C_u^{(1)}, C_{v-com}^{(T)}) \\ \vdots & \ddots & \vdots \\ \sum_r r P(r|C_u^{(K)}, C_{v-com}^{(1)}) & \dots & \sum_r r P(r|C_u^{(K)}, C_{v-com}^{(T)}) \end{bmatrix} \quad (4)$$

其中, $S_{com} \in \mathbb{R}^{K \times T}$ 中的元素 $S_{com}(i, j)$ 表示 Z 个领域中的用户聚类 i 对跨域信息对象的共性聚类 j 的评分期望值。然后,用函数 $f_{Re}(u, v) = p_u S_{com} p_{v-com}^T$ 表示用户 u 对信息对象 v 的基于跨领域共性特征的评分值。将式(1)、式(2)和式(4)分别代入函数,可得到:

$$f_{Re}(u, v) = \sum_r r \sum_{k,t} P(r|C_u^{(k)}, C_{v-com}^{(t)}) P(C_u^{(k)}|u) P(C_{v-com}^{(t)}|v) \quad (5)$$

类似地,对于第 z 个领域来说,可以构建基于单领域个性特征的用户评分模式,表示为:

$$S_{spe}^{(z)} = \begin{bmatrix} \sum_r r P(r|C_u^{(1)}, C_{v-spe}^{(1)}) & \dots & \sum_r r P(r|C_u^{(1)}, C_{v-spe}^{(L_z)}) \\ \vdots & \ddots & \vdots \\ \sum_r r P(r|C_u^{(K)}, C_{v-spe}^{(1)}) & \dots & \sum_r r P(r|C_u^{(K)}, C_{v-spe}^{(L_z)}) \end{bmatrix} \quad (6)$$

其中, $S_{spe}^{(z)} \in \mathbb{R}^{K \times T}$ 中的元素 $S_{spe}^{(z)}(i, j)$ 表示第 z 个领域中的用户聚类 i 对本域信息对象的特定聚类 j 的评分期望值。然后,用函数 $f_{Re-spe}^{(z)}(u, v) = p_u S_{spe}^{(z)} p_{v-spe}^T$ 表示用户 u 对信息对象 v



的基于单领域特定特征的评分值,可进一步得到:

$$f_{R-spe}^{(z)}(u, v) = \sum_r r \sum_{k, l_z} P(r | C_u^{(k)}, C_{v-spe}^{l_z}) P(C_u^{(k)} | u) P(C_{v-spe}^{l_z} | v) \quad (7)$$

然后,通过共享跨域共性特征信息的方式来缓解特定目标领域的数据稀疏性问题,并利用各领域中用户评分的个性特征提高跨域信息推荐的准确度,可以得到跨领域中用户 u 对信息对象 v 的最终评分函数 $f_R(u, v)$, 表示为:

$$f_R(u, v) = W_1^{(z)} f_{R-com}(u, v) + W_2^{(z)} f_{R-spe}^{(z)}(u, v) \quad (8)$$

其中, $W_1^{(z)}$ 表示用户评分中跨领域共性特征的权重系数, $W_2^{(z)}$ 表示用户评分中第 z 个领域的个性特征的权重系数,且 $W_1^{(z)} + W_2^{(z)} = 1$ 。

基于潜在因子模型的跨领域信息推荐(PCLF)算法示例如图1所示。算法中各潜在因子及相关参数可以通过期望最大化(expectation-maximization, EM)方法学习获得。

4 实验结果及分析

4.1 实验数据

选取多个领域的用户评分数据作为实验数据集^[14],包括:MovieLens 数据集,含有 943 个用户、1 682 部电影和超过 100 000 个评分数据(取值范围为 1~5 分);EachMovie 数据集,含有 72 916 个用户、6 214 部电影和超过 200 多万评分数据(取值范围 1~6 分,本文将其映射到 1~5 分);Book-Crossing 数据集,含有 278 858 个用户、271 379 本书和超过 100 多万评分数据(取值范围 0~9 分,本文将其映射到 1~5 分)。实验中将上述数据集中 70% 的用户及其评分数据作为训练样本,剩余部分作为测试样本。针对跨领域的实验,设定 Book-Crossing 与 EachMovie、Book-Crossing

与 MovieLens 和 MovieLens 与 EachMovie 数据集作为 3 种测试环境,验证跨领域信息推荐算法,在每个测试环境下重复实验 10 次并计算其平均值作为实验结果。

本文选择了其他 3 种相关算法与 PCLF 算法进行比较:NMF(nonnegative matrix factorization,非负矩阵分解)算法^[15],作为单领域推荐的基准算法来测试潜在因子模型在每个领域中的推荐性能;FMM(flexible mixture model)算法^[16],作为单域推荐的概率方法来测试基于聚类的概率算法在单领域中的推荐性能;RMGM(rating-matrix generative model,评分矩阵生成模型)算法^[17],作为当前跨领域推荐的最好算法来测试跨领域的推荐性能。所有的实验测试均采用 MAE(mean absolute error,平均绝对误差)作为衡量标准,MAE 的值越小,算法的性能越好。

4.2 实验结果及分析

PCLF 算法的参数设为 $K=20$ 、 $L_1=L_2=15$ 、 $T=10$, Given5、Given10 和 Given15 分别表示使用数据集进行训练时,对每个用户的评分数据只取前 5 个、10 个和 15 个数值的情况,其余的数据用来测试。

各种算法针对 Book-Crossing 和 EachMovie 数据集在不同实验设置下的 MAE 见表 1。从表 1 可以看出,在进行实验的 4 种算法中,PCLF 算法的 MAE 测试效果最好,而且,跨领域的推荐算法(PLCF 和 RMGM)比针对单领域的推荐算法(NMF 和 FMM)的性能好,这表明通过综合分析不同领域内用户对信息对象的评价数据,可以更好地找到跨领域用户之间的评分模式,从而缓解单个领域内数据的稀疏性问题。其中,RMGM 算法可以通过潜在因子模型抽取不同域中某一用户聚类集合对某一信息对象聚类集合评分模式的跨领域共性特征从而实现跨领域信息推荐,

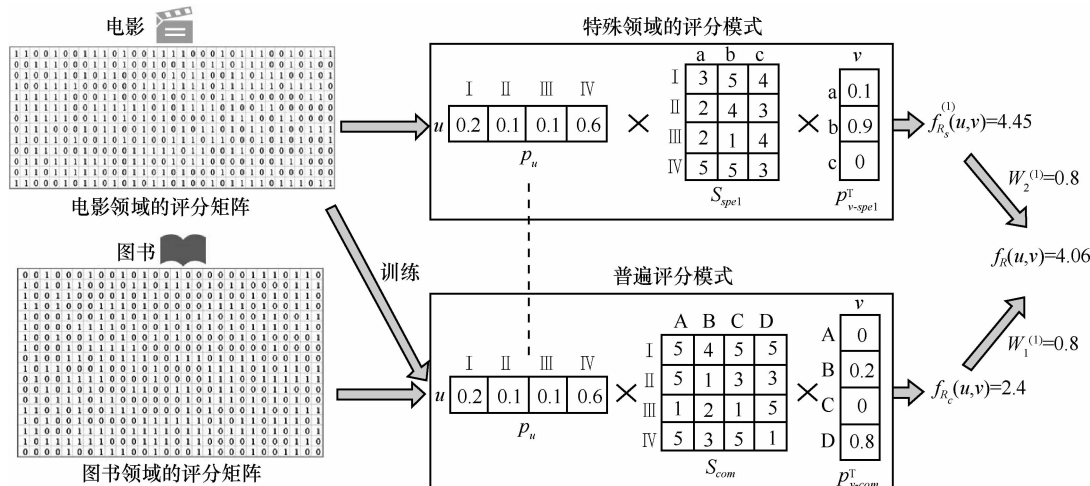


图1 基于潜在因子模型的跨领域信息推荐算法示例

而 PCLF 算法不仅可以提取出跨领域共性特征,还可以分析出每个域内用户聚类对信息对象聚类的评价模式的个性特征,进一步提高了跨域信息推荐的准确度。

表 1 各种算法针对 Book-Crossing 和 EachMovie 数据集在不同实验设置下的 MAE 对比

数据集	算法	Given5	Given10	Given15
Book-Crossing	NMF	0.657 5	0.637 5	0.632 1
	FMM	0.645 1	0.619 6	0.612 5
	RMGM	0.637 8	0.611 5	0.609 2
	PCLF	0.625 2	0.599 4	0.596 9
EachMovie	NMF	0.934 5	0.886 1	0.879 9
	FMM	0.913 2	0.883 1	0.877 1
	RMGM	0.902 1	0.874 3	0.863 7
	PCLF	0.883 8	0.867 7	0.853 3

各种算法针对 Book-Crossing 和 MovieLens 数据集、MovieLens 和 EachMovie 数据集在不同实验设置下的 MAE 见表 2 和表 3。表 2 和表 3 中的实验结果进一步验证了跨领域推荐算法比单领域推荐算法的性能好,也充分表明了不同领域之间的用户评分模式具有一定的关联性,通过提取跨领域用户评分模式的共性特征和个性特征可以缓解评分数据的稀疏性问题,也可以提高推荐的准确度。

表 2 各种算法针对 Book-Crossing 和 MovieLens 数据集在不同实验设置下的 MAE 对比

数据集	算法	Given5	Given10	Given15
Book-Crossing	NMF	0.657 5	0.637 5	0.630 1
	FMM	0.645 1	0.619 6	0.612 5
	RMGM	0.636 4	0.605 4	0.599 4
	PCLF	0.628 4	0.599 0	0.595 0
MovieLens	NMF	0.836 5	0.801 6	0.793 3
	FMM	0.805 2	0.783 8	0.772 1
	RMGM	0.803 9	0.776 9	0.769 2
	PCLF	0.792 2	0.772 9	0.765 3

表 3 各种算法针对 MovieLens 和 EachMovie 数据集在不同实验设置下的 MAE 对比

数据集	算法	Given5	Given10	Given15
MovieLens	NMF	0.836 5	0.801 6	0.793 3
	FMM	0.805 2	0.783 8	0.772 1
	RMGM	0.798 5	0.773 6	0.767 7
	PCLF	0.787 8	0.772 4	0.763 9
EachMovie	NMF	0.934 5	0.886 1	0.879 9
	FMM	0.913 2	0.883 1	0.877 1
	RMGM	0.901 7	0.873 3	0.860 1
	PCLF	0.882 0	0.865 8	0.849 8

5 结束语

针对跨领域信息推荐问题,本文通过综合分析不同领域内用户对信息对象的评价数据,基于潜在因子模型抽取了不同领域中某一用户聚类集合对某一信息对象聚类集合评分模式的跨领域共性特征和单领域个性特征,进而通过传递、共享跨领域共性特征信息的方式缓解了特定目标领域的数据稀疏性问题,提高了跨领域信息对象推荐结果的准确性和多样性,提升了跨领域信息资源的利用率。

参考文献

- Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering(TKDE)*, 2005,17(6): 734~749
- Koren Y. Collaborative filtering with temporal dynamics. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, 2009: 447~455
- Jordan M I, Blei D M, Ng A Y, *et al.* Latent dirichlet allocation. *Journal of Machine Learning Research*, 2003(3): 993~1022
- Hofmann T. Probabilistic latent semantic indexing. *Proceedings of the 21th Annual International SIGIR Conference on Research & Development in Information Retrieval*, Berkeley, CA, USA, 1999: 457~465
- Hoffman T. Collaborative filtering via gaussian probabilistic latent semantic analysis. *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Toronto, Canada, 2003: 259~266
- Bishop C M. *Pattern Recognition and Machine Learning*. New York: Springer-Verlag, 2006
- Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization. *Nature*, 1999, 401(6755): 788~791
- Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996(381): 607~609
- Gao S, Luo H, Chen D, *et al.* Cross-domain recommendation via cluster-level latent factor model. *Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Database*, Prague, Czech Republic, 2013: 161~176
- Pan W, Liu N N, Xiang E W, *et al.* Transfer learning to predict missing ratings via heterogeneous user feedbacks. *Proceedings of the 22th International Joint Conference on Artificial Intelligence*, Barcelona, Spain, 2011: 2318~2323

(2015188-6 接排在 2015179-6 后)

行预测,通过对预测值的修正,得到样本的整体分类准确率为 81%,证明本文中的模型对音乐风格的分类具有重要的应用价值。由于从网络采集到的语料有限,网络新词多变,使用情感语料库覆盖有限,使得神经网络的学习能力有一定限制,误判一些歌曲的音乐心情。如果有适合的网络情感新词发现算法,不断更新情感语料库,则可大大提高音乐心情的分类效果,一种新的识别情感生词算法,有待深入研究。目前标注的语料库普遍规模较小,涉及的领域和主题较少,大大降低了分类的效果。因此组织具有专业音乐知识的专家和计算机信息处理知识的专家共同制定规则,对领域和主题丰富的大规模语料库进行标注,建设一个标准的音乐风格语料库是具有广阔的发展前景。另外,此模型是以音乐心情的歌词为特征表示歌曲的,如果适当改变抽取规则针对主题去抽取关键词,则可以推广到以主题风格分类的音乐识别模型中。

参考文献

- 1 jsoup. <http://jsoup.org/>, 2015
- 2 ICTCLAS. <http://www.ictclas.org/>, 2015
- 3 潘文彬.基于情感词词典的中文句子情感倾向分析(硕士学位论文). 北京邮电大学, 2011
- 4 Lin P Y, Lin Z J, Kuang B Q, *et al.* A short chinese text incremental clustering algorithm based on weighted semantics and naive Bayes. *Journal of Computational Information Systems*, 2012, 8(10): 4257~4268
- 5 Hagan M T, Demuth H B, Beale M. *Neural Network Design*. PWS Pub Co, 2002
- 6 Mitchell T. *Machine Learning*. Boston: MIP Press, 2002
- 7 Baeza-Yates R, Ribeiro-Neto B. *Modern information retrieval*. Facet Publishing, 1999, 20(19): 305~306
- 8 Hu Y F, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. *Proceedings of IEEE 8th Conference on Data Mining, Pisa, Italy*, 2008: 263~272
- 9 Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model. *Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada*, 2009: 617~624
- 10 Singh A P, Gordon G J. Relational learning via collective matrix factorization. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA*, 2008: 650~658
- 11 Berkovsky S, Kuflik T, Ricci F. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, 2008, 18(3): 345~352
- 12 Zhang Y, Koren J. Efficient bayesian hierarchical user modeling for recommendation systems. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA*, 2007: 47~54
- 13 Lee D D, Seung H S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 2001, 13(3): 556~562
- 14 Luo S, Rong J. Flexible mixture model for collaborative filtering. *Proceedings of the 20th International Conference on Machine Learning, Washington DC, USA*, 2003(3): 704~711
- 15 张键锋,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。
- 16 王劲,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。
- 17 高升,男,博士,北京邮电大学副教授,主要研究方向为机器学习、信息推荐和社会网络分析等。
- 18 任思婷,女,北京邮电大学在读,主要研究方向为机器学习与信息推荐。
- 19 郭军,男,博士,北京邮电大学教授,主要研究方向为复杂网络与网络搜索等。

[作者简介]



张键锋,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。

王劲,男,广东省电信规划设计院有限公司工程师,主要从事无线、数据网络规划设计工作。

2015179-6

(上接 2015188-5)

- 11 Li B, Yang Q, Xue X. Transfer learning for collaborative filtering via a rating-matrix generative model. *Proceedings of the 26th International Conference on Machine Learning, Montreal, Canada*, 2009: 617~624
- 12 Singh A P, Gordon G J. Relational learning via collective matrix factorization. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA*, 2008: 650~658
- 13 Berkovsky S, Kuflik T, Ricci F. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction*, 2008, 18(3): 345~352
- 14 Zhang Y, Koren J. Efficient bayesian hierarchical user modeling for recommendation systems. *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA*, 2007: 47~54
- 15 Lee D D, Seung H S. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 2001, 13(3): 556~562
- 16 Luo S, Rong J. Flexible mixture model for collaborative filtering. *Proceedings of the 20th International Conference on Machine Learning, Washington DC, USA*, 2003(3): 704~711
- 17 Baeza-Yates R, Ribeiro-Neto B. *Modern information retrieval*. Facet Publishing, 1999, 20(19): 305~306
- 18 Hu Y F, Koren Y, Volinsky C. Collaborative filtering for implicit feedback datasets. *Proceedings of IEEE 8th Conference on Data Mining, Pisa, Italy*, 2008: 263~272

[作者简介]



高升,男,博士,北京邮电大学副教授,主要研究方向为机器学习、信息推荐和社会网络分析等。

任思婷,女,北京邮电大学在读,主要研究方向为机器学习与信息推荐。

郭军,男,博士,北京邮电大学教授,主要研究方向为复杂网络与网络搜索等。

2015188-6