计算机研究与发展 ISSN 1000-1239/CN 11-1777/TP
Journal of Computer Research and Development 45(9): 1463-1469, 2008

结合似然关系模型和用户等级的协同过滤推荐算法

高 滢 齐 红 刘 杰 刘大有
(吉林大学计算机科学与技术学院 长春 130012)
(吉林大学符号计算与知识工程教育部重点实验室 长春 130012)
(gying@jlu.edu.cn)

A Collaborative Filtering Recommendation Algorithm Combining Probabilistic Relational Models and User Grade

Gao Ying, Qi Hong, Liu Jie, and Liu Dayou
(*College of Computer Science and Technology, Jilin University, Changchun* 130012)
(*Ministry of Education, Key Laboratory of Symbolic Computation and Knowledge Engineering, Jilin University, Changchun* 130012)

Abstract Collaborative filtering is one of successful technologies for building recommender systems. Unfortunately, it suffers from sparsity and scalability problems. To address these problems, a collaborative filtering recommendation algorithm combining probabilistic relational models and user grade (PRM-UG-CF) is presented. PRM-UG-CF has primary two parts. First, a user grade function is defined, and user grade based collaborative filtering method is used, which can find neighbors for the target user only in his near grade, and the number of candidate neighbors can be controlled by a parameter, so recommendation efficiency is increased and it solves the scalability problem. Second, in order to use various kinds of information for recommendation, user grade based collaborative filtering method is combined with probabilistic relational models (PRM), thus it can integrate user information, item information and user-item rating data, and use adaptive strategies for different grade users, so recommendation quality is improved and it solves the sparsity problem. The experimental results on MovieLens data set show that the algorithm PRM-UG-CF has higher recommendation quality than a pure PRM-based or a pure collaborative filtering approach, and it also has much higher recommendation efficiency than a pure collaborative filtering approach.

Key words recommendation algorithm; collaborative filtering; probabilistic relational model; user grade; mean absolute error (MAE)

摘 要 针对传统协同过滤推荐算法的稀疏性、扩展性问题,提出了结合似然关系模型和用户等级的协同过滤推荐算法.首先,定义了用户等级函数,采用基于用户等级的协同过滤方法,在不影响推荐质量的前提下有效提高了推荐效率,从而解决扩展性问题;然后,将其与似然关系模型相结合,使之能够综合利用用户信息、项目信息、用户对项目的评分数据,对不同用户给出不同的推荐策略,从而解决稀疏性问题,提高推荐质量.在MovieLens数据集上的实验结果表明,该算法比单纯使用基于似然关系模型或传统协同过滤技术的推荐算法,不仅推荐质量有所提高,推荐速度比传统协同过滤算法明显加快.

关键词 推荐算法;协同过滤;似然关系模型;用户等级;平均绝对偏差

中图法分类号 TP391

收稿日期:2007-04-12;修回日期:2008-05-13

基金项目:国家自然科学基金重大项目(60496321);国家自然科学基金项目(60573073,60773099);国家"八六三"高技术研究发展计划基金项目(2006AA10Z245,2006AA10A309);吉林省科技发展计划基金项目(20030523);欧盟项目(TH/Asia Link/010(111084))

随着互联网的普及和电子商务的发展, 推荐系统逐渐成为电子商务 IT 技术的一个重要研究内容.

迄今为止, 研究者们已给出了许多适用于推荐系统的技术, 譬如协同过滤技术、Bayesian 网技术、聚类分析技术、关联规则技术、神经网络技术和图模型技术等.

协同过滤技术是目前比较成功的推荐技术[1]. 协同过滤的基本思想是基于目标用户最近邻居的评分数据, 预测目标用户对未评分项目的评分, 从而将预测评分最高的若干项目推荐给目标用户. 然而, 协同过滤推荐也存在一些局限性: 1)稀疏性, 即当用户评分数据较少时推荐质量较低, 特别是对于系统的新用户和新项目不能产生推荐; 2)扩展性, 即随着系统规模的扩大用户数目和项目数目逐渐增多, 推荐效率明显降低.

为了解决数据稀疏性问题, 提高协同过滤推荐的质量, 目前有些工作采用先将稀疏矩阵部分或全部填充[2-4], 然后再用传统协同过滤的方法; 还有一些工作把基于内容的过滤和协同过滤相结合, 同时考虑多种可用信息, 以提高推荐质量[4-5]. 实验表明, 这两类方法与传统协同过滤算法相比, 在推荐质量上都有所提高, 但从效率来看所需的时间更多.

为了解决系统扩展性问题, 提高协同过滤推荐的效率, 常用聚类分析的方法. 将最近邻居搜索限制在目标用户所在的类别中, 或者直接从聚类中心提取推荐结果, 该方法虽然能提高推荐速度, 但多数会降低推荐质量[1,6]; 赵亮等人通过矩阵奇异值分解来减少项目空间的维数[7], 以提高算法效率, 但降维通常会导致信息丢失[8].

1998 年, Koller 等人提出似然关系模型(probabilistic relational model, PRM)[9], 该模型将标准 Bayesian 网扩展到关系模式之上, 从而在类的层次上建立 Bayesian 网来表示类属性间的依赖关系. 关系模式 R 的似然关系模型对于 R 中的每个类 X 及其描述属性 A, 有: ①父结点集 $Pa(X.A)=\{U_1, \cdots, U_m\}$, 这里 $U_i(1\leqslant i\leqslant m)$ 为 $X.B$ 或 $X.\tau.B$, τ 为引用链(reference chains); ②条件概率分布 $P(X.A|Pa(X.A))$[10]. 2004 年, Newton 等人提出基于 PRM 的推荐技术[11]. 基于 PRM 的推荐所依据的信息不仅包括用户对项目的评分数据, 还包括用户信息、项目信息表; 一旦模型构造完毕, 其推荐效率很高; 不足之处是推荐质量有待进一步提高.

为了同时解决协同过滤的稀疏性、扩展性问题, 本文首先提出基于用户等级的协同过滤技术, 在对用户间相似性公式进行分析的基础上, 通过定义用户等级函数, 采用仅在用户等级的邻域内查找近邻的方法, 在不影响推荐质量的前提下, 大大提高推荐效率, 有效解决扩展性问题; 然后, 将其与基于似然关系模型的推荐技术相结合, 使其在推荐过程中能够综合利用多种类型的信息, 并根据用户等级计算结合权重, 从而解决数据稀疏性问题, 提高推荐质量.

1 结合似然关系模型和用户等级的协同过滤推荐算法

1.1 用户等级函数

用户对项目的评分数据是为用户推荐的主要依据, 该数据可以表示为 $m\times n$ 阶的矩阵 $\boldsymbol{R}$, m 代表用户数, n 代表项目数, 元素 R_{ij} 代表用户 i 对项目 j 的评分.

本文按照用户所评价的项目数量, 将其分为不同的等级, 并引入用户等级函数. 用户等级函数以用户所评价的项目数量为自变量, 函数值与自变量成正比, 反映用户与系统交互的程度. 另外, 为了反映出不同等级的用户数目的统计特性, 本文利用用户等级函数的增长率表示用户的分布信息.

为了给出用户等级函数的具体形式, 首先对数据集中的数据进行统计, 通过用户密度函数表示具有不同评价项目数量的用户数目的统计特性. 为了计算方便, 本文把用户密度函数、用户等级函数定义为连续函数.

定义 1. 设 X 为一个随机变量, 表示用户所评价的项目数量, 取值范围为 R^+. 任取 $x\in\mathrm{R}^+$, 则称函数 $f(x)$ 为 X 的用户密度函数, 表示具有不同评价项目数量的用户数目的统计特性.

定理 1. 设 x 是随机变量 X 的某一具体取值, δ 为某一正数, 且满足 $x-\delta>0$, 则 x 的 δ 邻域内的用户数 $N(x,\delta)=\int_{x-\delta}^{x+\delta} f(t)\mathrm{d}t$. 当 $\delta\ll x$ 时, $N(x,\delta)\approx 2\delta\times f(x)$.

证明. 根据用户密度函数的定义, $f(x)$ 表示所评价的项目数量为 x 的用户数目. δ 为正数, 且 $x-\delta>0$, 则所评价的项目数量在 $[x-\delta, x+\delta]$ 区间内的用户数 $N(x,\delta)$ 显然为 $\int_{x-\delta}^{x+\delta} f(t)\mathrm{d}t$.

当 $\delta\ll x$ 时, 对于 $t\in[x-\delta, x+\delta]$, 有 $f(t)\approx f(x)$, 所以 $N(x,\delta)\approx 2\delta\times f(x)$. 证毕.

定义 **2**. 设用户等级域 $M=\{r \mid 0\leqslant r\leqslant 1, r\in R\}$, 随机变量 X 表示用户所评价的项目数量, X 的取值范围为 R^+, 则称映射 $G: R^+\rightarrow M$ 为用户等级函数, 其满足:

① 是关于自变量 X 的增函数;

② 函数增长率与用户密度函数成正比, 即 $G'(x)\propto f(x)$.

按照用户等级函数的定义, 若某用户当前所评价的项目数量为 x, 则其等级为 $G(x)$.

定理 **2**. 设任一用户所评价的项目数量为 x, 则对某较小常数 α, 等级在 $G(x)$的 α 邻域内的用户数仅与 α 相关, 且与 α 成正比.

证明. 当 α 较小时, 对于所评价的项目数量为 x 的用户, 等级 $G(x)$的 α 邻域区间大小为 2α. 根据微分的性质, 有 $\triangle y\approx dy=G'(x)\times dx$, 所以, $2\alpha\approx G'(x)\times 2\triangle x$, 这里 $2\triangle x$ 为所评价的项目数量 x 的邻域区间大小.

根据用户等级函数的定义, $G'(x)\propto f(x)$, 不妨设 $G'(x)=b\times f(x)$, 这里 b 为常数.

根据定理 1, 当 $\triangle x$ 较小时, 该区间内的用户数 $N(x, \triangle x)\approx 2\triangle x\times f(x)$. 所以, $N(x, \triangle x)\approx 2\triangle x\times G'(x)/b\approx 2\alpha/b$.

由此可知, 对于所评价的项目数量为 x 的用户, $G(x)$的 α 邻域区间内的用户数与 x 无关, 仅与 α 相关, 且与 α 成正比. 证毕.

定理 **3**. 等级为 r 的用户数目在 r 的取值区间 [0, 1] 上近似服从均匀分布.

证明. 由定理 2 可知, 对某较小常数 α, 等级 r 的 α 邻域内的用户数仅与 α 相关, 且与 α 成正比.

所以, 等级为 r 的用户数目在 r 的取值区间[0, 1] 上近似服从均匀分布. 证毕.

1.2　基于用户等级的协同过滤

协同过滤推荐过程可分为两步[12]:

1) 发现与目标用户评分最相近的若干邻居. 传统找近邻方法是计算目标用户与其他每个用户的相似度, 并把相似度最高的若干用户作为近邻. 修正的余弦相似度是衡量用户间相似性的较好方法[13]:

$$sim_{uv}=\frac{\sum_{i\in I_{uv}}\left(R_{ui}-R_u\right)\times\left(R_{vi}-R_v\right)}{\sqrt{\sum_{i\in I_u}\left(R_{ui}-R_u\right)^2}\sqrt{\sum_{i\in I_v}\left(R_{vi}-R_v\right)^2}}, \tag{1}$$

其中, sim_{uv}表示用户 u 与用户 v 的相似度, I_u, I_v 分别为用户 u, v 已评分项目集, I_{uv} 为共同评分项目集, R_{ui}, R_{vi} 为用户 u, v 对项目 i 的评分, R_u, R_v 为用户 u, v 的平均评分.

2) 依据近邻评分数据, 预测目标用户对未评分项目的评分, 并将预测评分最高的若干项目推荐给目标用户. 预测方法如下[1]:

$$P_{ui}^{NBS}=R_u+\frac{\sum_{v\in NBS}sim_{uv}\times(R_{vi}-R_v)}{\sum_{v\in NBS}(|sim_{uv}|)}, \tag{2}$$

P_{ui}^{NBS} 表示用户 u 对项目 i 的预测评分, NBS 为用户 u 的近邻集.

通过对式(1)的分析可以看出, 两用户的相似度大小与其所评价的项目集相关, 其中包括项目集的大小、项目集所包含的项目及其评分值. 对于相似度高的两用户, 其所评价的项目数量相差一定不会太大, 因为所评价的项目数量相差悬殊的用户间的相似度不可能太大; 对于所评价的项目数量相近的用户, 相似度越高表明用户共同评分的项目数所占的比例越大, 且评分值也越相似.

因此, 本文提出基于用户等级的协同过滤技术, 为用户找近邻时, 首先把候选近邻的范围限定到与该用户所评价的项目数量相近的用户范围内, 然后在该候选近邻集内, 采用用户相似度计算公式, 进一步查找近邻. 根据上节对用户等级函数的定义和定理 3 可知, 可以把候选近邻的范围限定到用户等级的某邻域区间内, 并通过调节邻域区间的大小调节候选近邻集的用户数目. 这样, 一方面, 由于缩小查找范围提高了查找效率, 另一方面, 该区间内的用户所评价的项目数量接近, 在该区间内查找近邻, 其准确性并未受到影响.

为了加快推荐速度, 本文将式(2)中的项目 i 限定到如下集合:

$$I_{NBS}=\bigcup_{i\in NBS}I_i. \tag{3}$$

1.3　结合似然关系模型的推荐

协同过滤推荐技术的推荐质量依赖于用户所评价的项目数量. 用户所评价的项目数量越多评分数据越能准确反映用户的兴趣爱好, 推荐质量越高. 为了解决低等级用户推荐质量低的问题, 本文将基于用户等级的协同过滤与 PRM 相结合, 使其在推荐过程中能够同时使用多种信息.

结合之后, 用户 u 对项目 i 的最终预测评分取两种方法预测评分的加权平均值, 即

$$P_{ui}=G_u\times P_{ui}^{NBS}+(1-G_u)\times P_{ui}^{PRM}, \tag{4}$$

其中, G_u 为用户 u 的等级, P_{ui}^{NBS} 为基于协同过滤的预测评分, P_{ui}^{PRM} 为基于 PRM 的预测评分. 用户等级

越低协同过滤的推荐质量越差, 因此模型信息所占的比重越大, 当用户等级值为 0 时, 预测结果完全由模型信息确定. 反之, 随着用户等级的增加, 协同过滤的推荐质量提高, 所占的比重也逐渐增大, 当用户等级值为 1 时, 预测结果完全由协同过滤算法确定.

基于 PRM, 用户 u 对项目 i 的预测评分[11]:

$$P_{ui}^{PRM} = E(R.rating \mid Pa(R.rating)), \quad (5)$$

其中, $R.rating$ 为评分结点, $Pa(R.rating)$为其父结点集, 父结点集与用户和项目的属性相关. 预测评分 P_{ui}^{PRM} 是根据父结点的条件概率表计算的用户 u 对项目 i 的期望评分. 为了加快推荐速度, 本文将预测评分的项目 i 限定到用户 u 期望评分最高的项目类, 即

$$I_{PRM} = argmax\ E(R.rating \mid Pa(R.rating)). \quad (6)$$

因此, 结合似然关系模型与用户等级的协同过滤推荐算法, 候选推荐项目集为

$$I_{NBS} \cup I_{PRM}. \quad (7)$$

1.4 算法及时间复杂度分析

如算法 1 所示, 结合似然关系模型和用户等级的协同过滤推荐算法(PRM-UG-CF), 首先根据用户项目评分矩阵 $\boldsymbol{R}$ 统计每个用户所评价的项目数量, 并生成用户等级函数, 由此求得目标用户 u 的等级 G_u. 然后, 根据基于用户等级的协同过滤, 在 G_u 的 α 邻域内求 u 的近邻集, 并利用式(3), 求基于近邻的候选推荐项目集 I_{NBS}. 再根据用户信息表 U、项目信息表 I、用户项目评分矩阵 $\boldsymbol{R}$, 创建 PRM 模型, 并使用式(6)求基于 PRM 的候选推荐项目集 I_{PRM}. 最后, 对 $I_{NBS} \cup I_{PRM}$ 中用户未评分项目, 使用式(4)进行评分预测, 并将预测评分最高的若干项目推荐给目标用户.

如果把算法中第①, ⑤步作为预处理, 影响算法效率的主要步骤是第③及第⑦ ~ ⑭步. 在第③步中, 传统协同过滤算法的时间复杂度为 $\Theta(mn)$, 而基于用户等级的协同过滤算法复杂度有所降低, 降低幅度与 α 相关. 第⑦ ~ ⑭步的时间复杂度与候选推荐项目集的大小相关, 最坏情况为 $O(n)$. 可见, 影响算法时间复杂度的主要步骤为第③步, 即为用户查找近邻的操作.

算法 1. PRM-UG-CF(U, I, $\boldsymbol{R}$, u. $RecList$).

输入: 用户信息表 U、项目信息表 I、评分矩阵 $\boldsymbol{R}$、目标用户 u.

输出: 用户 u 的推荐项目列表 $RecList$.

① $CreateUserGradeFunction(\boldsymbol{R}.G)$ /*生成用户等级函数 G */

② $GetGrade(G, u.G_u)$ /*获得用户 u 的等级值 G_u */

③ $FindNeighbors(u, G_u, \alpha.NBS)$ /*在用户等级值的 α 邻域内查找用户 u 的近邻集 NBS */

④ $GenerateCandidateItemsNBS(NBS, \boldsymbol{R}.I_{NBS})$ /*将 NBS 评分项目集作为候选推荐项目集 I_{NBS} */

⑤ $CreatePRM(\boldsymbol{R}, U, I.PRM)$ /*创建 PRM 模型 */

⑥ $GenerateCandidateItemsPRM(PRM.I_{PRM})$ /*根据 PRM 求候选推荐项目集 I_{PRM} */

⑦ for each $i \in I_{NBS} \cup I_{PRM}$

⑧ if R_{ui} = null then

⑨ $NBSBasedPredict(u, i, NBS.P_1)$ /*对未评分项目采用基于用户等级的协同过滤, 计算预测评分 P_1 */

⑩ $PRMBasedPredict(u, i, PRM.P_2)$ /*对未评分项目, 基于似然关系模型计算用户的预测评分 P_2 */

⑪ $P_{ui} \leftarrow G_u \times P_1 + (1 - G_u) \times P_2$ /*以用户等级为权值计算加权预测评分 */

⑫ endif

⑬ endfor

⑭ $RecList \leftarrow FirstTopN(P_{ui})$ /*将预测评分最高的 N 项作为推荐列表返回 */

2 实验结果

2.1 数据集

本文采用 MovieLens 站点提供的数据集(http://www.grouplens.org/data/million/). 该版本数据集共含有 6040 个用户对 3883 个电影的 1000209 个评价记录评分值采用 5 分制.

2.2 度量标准

本文选用平均绝对偏差(MAE)作为评价推荐系统推荐质量的度量标准, MAE 定义[13]为

$$MAE = \frac{\sum_{i=1}^{N} |p_i - q_i|}{N}, \quad (8)$$

N 为测试集大小, p_i 为预测评分, q_i 为实际评分.

MAE 越小推荐质量越高.

2.3　实验结果及分析

2.3.1　用户等级函数

统计数据集中具有不同评价项目数量的用户数目所得结果如图 1 所示. 采用最小二乘法对其进行多项式函数拟合,并对拟合结果积分,再将其线性变换到[0,1]区间,由此求得用户等级函数,结果如图 2 所示,具体形式为

$$y=\begin{cases}0, & x<20,\\ b\times(a_1/3\times x^3+a_2/2\times x^2+\\ \quad a_3\times x+a_4), & 20\leqslant x\leqslant 320,\\ 1, & x>320,\end{cases}$$

其中,$b=0.0149$, $a_1=1.5850e-005$, $a_2=-0.0075$, $a_3=0.9213$, $a_4=-16.96$.

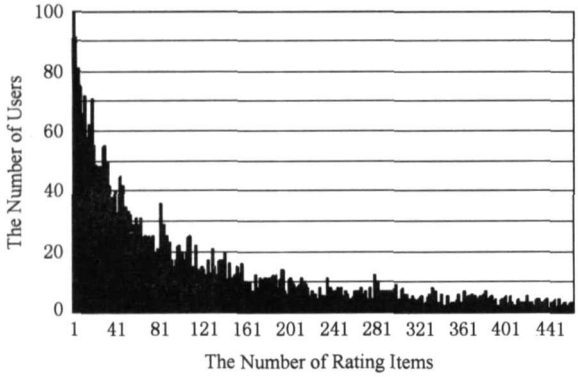

Fig. 1　Statistical diagram of user numbers with different numbers of rating items.

图 1　具有不同评价项目数量的用户数统计图

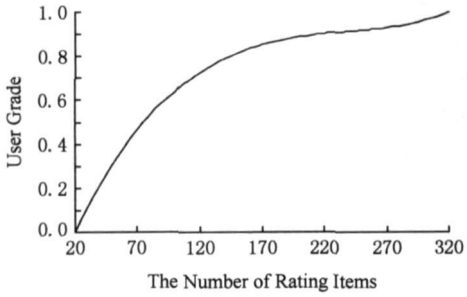

Fig. 2　User grade function.

图 2　用户等级函数

2.3.2　建立 PRM

本文采用文献[10]中的方法为用户信息、电影信息及用户对电影的评价信息建立模型,图 3 为标准 PRM,图 4 为针对某具体用户 *uid* 和具体电影 *mid* 的 Bayesian 网,其中的参数省略.

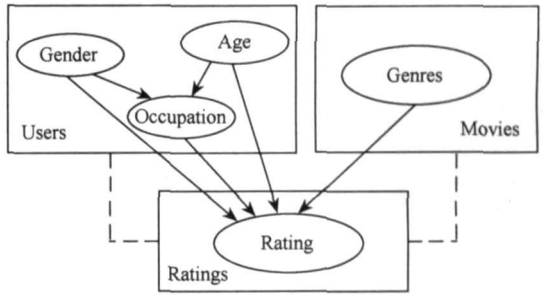

Fig. 3　Standard PRM.

图 3　标准 PRM

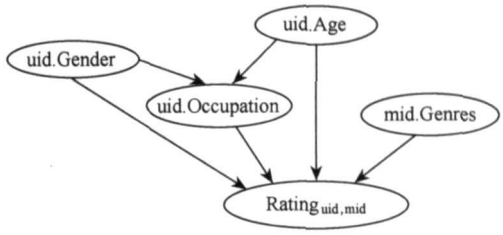

Fig. 4　Ground Bayesian network.

图 4　实例层次的 Bayesian 网

2.3.3　推荐结果

1)推荐效率

表 1 为基于用户等级的协同过滤算法(UG-CF)和传统协同过滤算法(CF)候选近邻数随 α 值的变化情况. 在 Pentium Ⅳ 1.7GHz, DDR512 MB 内存的PC 机上,用 Java 语言做实验,为不同等级的用户查找近邻所需时间见表 2,表中时间单位为秒,实验中每个用户的近邻数为 10,最终我们将 α 取值为 0.03.

Table 1　Comparison of the Number of Candidate Neighbors

表 1　候选近邻数比较

Algorithm	α				
	0.01	0.02	0.03	0.05	0.08
UG-CF	83	162	241	434	689
CF	6040	6040	6040	6040	6040

Table 2　Time of Finding Neighbors

表 2　查找近邻时间　s

Algorithm	User Grade				
	0.1	0.3	0.5	0.7	0.9
UG-CF	6	11	14	21	38
CF	141	215	317	495	984

2)推荐质量

在实验中,将每个用户的评分信息分成相等的

5份,采用交叉验证的方法,每次取一份为测试集,其余4份组合在一起为训练集,进行5次实验.图5反映了UG-CF和CF的*MAE*随用户等级的变化情况,将其与PRM结合之后各种算法平均*MAE*如图6所示.

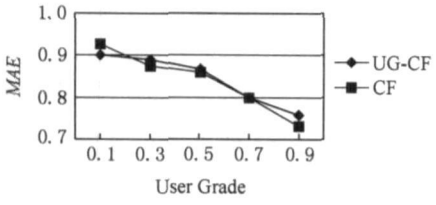

Fig. 5 *MAE* of different grade users.

图5 不同等级用户的*MAE*

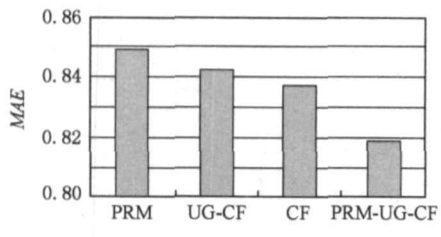

Fig. 6 *MAE* of different algorithms.

图6 各种算法的*MAE*

A2.3.4 实验结果分析

表1及表2中的实验数据表明,UG-CF和CF相比,由于候选近邻数目减少,算法的效率明显提高,而对推荐质量的影响较小,$\alpha=0.03$时,UG-CF比CF的*MAE*仅增加0.06%.由图5可知,随着用户等级的增加,UG-CF和CF算法的*MAE*明显下降,说明用户所评价的项目数量越多推荐质量越好,反之,当用户等级较低时误差较大(*MAE*超过0.9).由图6可知,基于PRM的算法误差最大,但将其与协同过滤算法相结合推荐质量明显提高.

3 结 论

为了解决传统协同过滤推荐的稀疏性和扩展性问题,本文提出了结合似然关系模型和用户等级的协同过滤推荐算法PRM-UG-CF.该算法可以综合利用用户信息、项目信息和用户对项目的评分数据,并根据用户等级值确定各种信息在推荐过程中所占的比重,提高了推荐质量;基于用户等级的协同过滤技术,仅在用户等级的邻域内查找近邻的过程有利于推荐效率的提高.理论分析及实验结果表明,PRM-UG-CF比单纯使用基于似然关系模型或传统协同过滤技术的推荐算法,不仅推荐质量有所提高,推荐速度比传统协同过滤算法明显加快.

参 考 文 献

[1] Breese J, Heckerman D, Kadie C. Empirical analysis of predictive algorithms for collaborative filtering[C] // Proc of the 14th Conf on Uncertainty in Artificial Intelligence. San Francisco: Morgan Kaufmann, 1998: 43-52

[2] Deng Ailin, Zhu Yangyong, Shi Baile. A collaborative filtering recommendation algorithm based on item rating prediction[J]. Journal of Software, 2003, 14(9): 1621-1628 (in Chinese)
(邓爱林, 朱扬勇, 施伯乐. 基于项目评分预测的协同过滤推荐算法[J]. 软件学报, 2003, 14(9): 1621-1628)

[3] Zhou Junfeng, Tang Xian, Guo Jingfeng. An optimized collaborative filtering recommendation algorithm[J]. Journal of Computer Research and Development, 2004, 41(10): 1842-1847 (in Chinese)
(周军锋, 汤显, 郭景峰. 一种优化的协同过滤算法[J]. 计算机研究与发展, 2004, 41(10): 1842-1847)

[4] Meville P, Mooney R J, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C] // Proc of the 18th National Conf on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 2002: 187-192

[5] Li Q, Kim B M. An approach for combining content-based and collaborative filters[C] // Proc of the 6th Int Workshop on Information Retrieval with Asian Language. Morristown. NJ, USA: Association for Computational Linguistics, 2003: 17-24

[6] Mobasher B, Dai H, Luo T, *et al*. Integrating web usage and content mining for more effective personalization[C] // Proc of the Int Conf on the E-Commerce and Web Technologies. Berlin: Springer, 2000: 165-176

[7] Zhao Liang, Hu Naijing, Zhang Shouzhi. Algorithm design for personalization recommendation systems[J]. Journal of Computer Research and Development, 2002, 39(8): 986-991 (in Chinese)
(赵亮, 胡乃静, 张守志. 个性化推荐算法设计. 计算机研究与发展, 2002, 39(8): 986-991)

[8] Aggarwal C C. On the effects of dimensionality reduction on high dimensional similarity search[C] // Proc of the 20th ACM SIGMOD-SIGACT-SIGART Symp on Principles of Database Systems. New York: ACM Press, 2001: 256-266

[9] Koller D, Pfeffer A. Probabilistic frame-based systems[C] // Proc of the 15th National Conf on Artificial Intelligence. Menlo Park, CA, USA: AAAI Press, 1998: 580-587

[10] Friedman N, Getoor L, Koller D, *et al*. Learning probabilistic relational models[C] // Proc of the 16th Int Join Conf on Artificial Intelligence. San Francisco: Morgan Kaufmann, 1999: 1300-1309

[11] Newton J, Greiner R. Hierarchical probabilistic relational models for collaborative filtering[C] // Proc of the ICML 2004 Workshop on Statistical Relational Learning and its Connections to Other Fields. Banff, Canada: IMLS, 2004: 82-87

[12] Herlocker J L, Konstan J A, Borchers A, *et al*. An algorithmic framework for performing collaborative filtering[C] // Proc of the 22nd Annual Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM Press, 1999: 230-237

[13] Sarwar B, Karypis G, Konstan J, *et al*. Item-based collaborative filtering recommendation algorithms[C] // Proc of the 10th Int Conf on World Wide Web. New York: ACM Press, 2001: 285-295

Gao Ying, born in 1978. Ph. D. candidate and lecturer. Her main research interests include data mining and statistical relational learning.

高 滢, 1978 年生, 博士研究生, 讲师, 主要研究方向为数据挖掘、统计关系学习等.

Qi Hong, born in 1970. Ph. D. and associate professor. Her main research interests include data mining and statistical relational learning.

齐 红, 1970 年生, 博士, 副教授, 主要研究方向为数据挖掘、统计关系学习等.

Liu Jie, born in 1973. Ph. D. and lecturer. Her main research interests include data mining and pattern recognition.

刘 杰, 1973 年生, 博士, 讲师, 主要研究方向为数据挖掘、模式识别等.

Liu Dayou, born in 1942. Professor and Ph. D. supervisor. His main research interests include knowledge engineering and expert system, spatio-temporal reasoning and geographical information system, data mining, and multi agent system.

刘大有, 1942 年生, 教授, 博士生导师, 主要研究方向为知识工程与专家系统、时空推理与地理信息系统、数据挖掘、多 Agent 系统等(dyliu@jlu.edu.cn).

Research Background

This research is supported by NSFC Major Research Program under grant No. 60496321; the National Natural Science Foundation of China under grant Nos. 60773099 amd 60573073; the National High-Tech Research and Development Plan of China under grant Nos. 2006AA10Z245 and 2006AA10A309; the Major Program of Science and Technology Development Plan of Jilin Province under grant No. 20020303; the Science and Technology Development Plan of Jilin Province under grant No. 20030523; and European Commission under grant No. TH/Asia Link/010 (111084). Recommender systems help overcome information overload by providing personalized suggestions. Collaborative filtering is a very successful technique for building recommender systems, and it is the technique of using peer opinions to predict the interests of others. A target user is matched against the database to discover neighbors, who have historically similar interests to the target user. Items that neighbors like are then recommended to the target user. Although collaborative filtering has been successfully used in both research and practice, it suffers from sparsity and scalability problems. To address these problems, we present a collaborative filtering recommendation algorithm combining probabilistic relational models and user grade (PRM-UG-CF). PRM-UG-CF can integrate item information, user information and user-item rating data, and find neighbors efficiently, and it solves the sparsity and scalability problems simultaneously.