多标签学习在智能推荐中的研究与应用

朱峙成[1] 刘佳玮[1,2] 阎少宏[1,2]
(河北省数据科学与应用重点实验室 河北 唐山 063210)[1]
(华北理工大学数学建模创新实验室 河北 唐山 063210)[2]

摘 要 传统的智能推荐中运用了协同过滤算法,但是它并不能很好地处理用户的评分信息,推荐的质量受存在的数据稀疏性、极端数据的影响。对此,将推荐问题转换为多标签学习问题,文中提出了一种基于 HMM 模型和用户画像的完备智能推荐系统。首先设立不同的数据处理机制来提高模型的泛化能力,其次为了解决数据稀疏问题,提出反马尔科夫性改进 HMM 模型,最终构建用户画像对 HMM 模型的学习经验得到的结果进行筛选,得到最终的推荐服务。实验结果表明,在智能推荐问题中多标签学习有效地提高了推荐准确性和推荐效率。

关键词 智能推荐系统,多标签学习,数据稀疏,用户画像
中图法分类号 TP3-05 **文献标识码** A

Research and Application of Multi-label Learning in Intelligent Recommendation

ZHU Zhi-cheng[1] LIU Jia-wei[1,2] YAN Shao-hong[1,2]
(Hebei Key Laboratory of Data Science and Application, Tangshan, Hebei 063210, China)[1]
(Mathematical Modeling Innovation Laboratory, North China University of Science and Technology, Tangshan, Hebei 063210, China)[2]

Abstract Collaborative filtering algorithm is used in traditional intelligent recommendation, but it can't deal with user's rating information well. The data sparsity and extreme data influence the quality of recommendation. Therefore, the recommendation problem is transformed into a multi-label learning problem, and a complete intelligent recommendation system based on HMM model and user portrait was proposed in this paper. Firstly, different data processing mechanisms are set up to improve the generalization ability of the algorithm. Secondly, an improved HMM model with anti-Markov property is proposed to solve the problem of data sparsity. Finally, a user portrait is constructed to screen the learning experience of the HMM model and get the final recommendation service. Experimental results show that multi-label learning can effectively improve the accuracy and efficiency of intelligent recommendation.

Keywords Intelligent recommendation system, Multi-label learning, Data sparsity, User portrait

在互联网时代的背景下,各类信息量的日益增加使得用户难以在海量数据下找到符合自己需求的服务。在此背景下,推荐系统应运而生[1],提供一种能够在大量数据中推荐符合用户需求的服务。推荐系统首先是分析大量用户的历史数据等信息记录,然后建立模型分析与当前待推荐用户属性相似的特点,从而找到与用户需求相匹配的服务进行推荐,搭建智能推荐系统。一般来说,推荐技术中协同过滤推荐方法是推荐系统中较为成功和重要的方法[2]。然而,协同过滤推荐方法往往面临数据稀疏的问题[3],而且当数据集中含有不同类型的数据时,推荐效果往往较差,用户得不到满意的结果。

目前对于多标签学习推荐的研究大多是独立进行的[4]。事实上,在商品推荐中,推荐系统通常给一个待推荐对象一系列可能的商品服务。如果我们把这个被推荐商品的对象看作一个实例,推荐的一系列商品服务看作多个标签,那么智能推荐问题则可以转化为多标签学习问题。因多标签学习是解决信息过载的有效方法,因此,本文借鉴多标签学习方法讨论如何在海量数据中训练标签,产生学习经验,并利用学得的经验进行推荐。

概率图模型是一类基于概率相关关系的模型[5]。概率图模型拥有着优良的性质,它提供了一种简单的可视化概率模型的方法,用于表示复杂的推理和学习运算,简化数学表达。基于概率推理机制,度量用户或标签之间的相似性,即可获得多标签学习的经验;同时利用用户画像对推荐结果进行筛选,从而提供个性化服务,进一步满足用户的需求。

因此,本文在传统协同过滤算法的基础上同时借助多标签学习的思想提出了一种基于概率图的智能推荐方法。其采用概率图的方式考虑了相似用户属性,同时权衡了自身的因素,并且有效地解决了数据稀疏问题,并根据用户信息构建用户画像[6]对推荐特征进行有效筛选,提高了推荐服务的质量。

1 多标签学习模型的建立

标签具有多样性,标签类型大致可以分为 3 类。"0-1"数据推荐、"评分"数据推荐、"任意"数据推荐。在实际应用中,"任意"数据应用范围最广,应用场景最大,其他两种类型数据可以看作"任意"数据的特殊情况。

鉴于此,本文计算用户相似性时考虑服务的自身质量属

本文受到国家青年基金资助项目(11301120),河北省青年科学基金资助项目(A2015209189),唐山市数据科学重点实验室基金资助项目,河北省数据科学与应用重点实验室基金项目资助。

朱峙成(1998—),男,主要研究方向为智能计算、人工智能;**阎少宏**(1977—),男,硕士,副教授,硕士生导师,主要研究方向为深度学习、机器学习,E-mail:shaohong@ncst.edu.cn(通信作者)。

性,以及处理数据过程中可能出现的稀疏矩阵问题,提出了一种新式智能推荐算法。首先对问题进行转化,将商品属性看作标签,着重于标签间的数据挖掘,利用历史信息记录进行计算,将多标签学习与概率图模型相结合,在筛选用户画像后求解得到待推荐的标签。

1.1 基于概率图的智能推荐

物理世界中往往是表象,而表象背后的隐藏状态是不可见的,我们需要通过观察大量的表象来总结规律以认识事物的本质[7]。即:标签数据背后蕴含着用户喜好等隐藏状态,需要训练历史数据集并学习经验。为了利用学习到的经验达到智能推荐的目的,本文建立 HMM(Hidden Markov Model)模型[8]来解决上述需求。

1.1.1 HMM 模型的性质

(1)概率不动性

$P(X_{i+1}|X_i)=P(X_{j+1}|X_j)$,$\forall i,j$。即状态概率转移矩阵与时间无关。考虑到实际背景,此性质的应用范围改动为:对与大多数的用户而言,如果喜欢标签 L_1,那么在喜欢标签 L_1 之后喜欢标签 L_2 的概率相同,反之亦然。

(2)反马尔可夫性

马尔可夫性的定义为当前状态只与前一个状态有关,即:

$$P(X_{n+1}|X_0,\cdots,X_n)=P(X_{n+1}|X_n) \tag{1}$$

考虑到转化问题的需求,提出适合实际情况的反马尔可夫性,此性质改进为:对用户进行推荐时,之前的状态 $(X_0,\cdots,X_n)$ 的含义为基于大数据背景下学习得到的经验概率,当数据量越大时,经验越稳定。因此,对与待推荐的用户而言,推荐的结果状态会与之前使用标签的状态存在一定的关系。

1.1.2 HMM 模型的建立

(1)定义生成概率矩阵 $b_j(k)$。其中 $b_j(k)$ 是在时刻 t 处于状态 S 的条件下生成 V 的概率,即 $b_j(k)=P(O_t=V_k|i_t=S_j)j,k\in(1,n)$,$n$ 是标签状态总数。根据生成概率矩阵的定义,可知考虑了时间 t 的因素。但在实际分析问题时,考虑时间 t 是很繁琐的,甚至无从得知之前时间段内的状态。基于 HMM 性质(1)我们做出简化,即在任何时刻,生成概率矩阵随数据集的增加而趋于逐渐稳定,而忽略与时间相关的变化。

(2)计算状态转移概率矩阵 $A=a_{ij}$。其中 $a_{ij}=P(i_{t+1}=S_j|i_t=s_i)$,矩阵的含义为在时刻处于状态 S_i 的条件下,在下一时刻转移到状态 S_j 的概率。则此矩阵在本次应用中的含义为:对于用户 U 而言,喜欢标签 L_1 且喜欢标签 L_2 的概率。

(3)输入初始向量 π。π 的含义为初始状态概率向量,即在最初始的情况下,用户对每个标签的喜好程度,为公平起见,假设对每个标签的喜好程度相同,即 $\pi=(1/n,1/n,\cdots,1/n)$。至此,隐马尔可夫模型基本元素定义完毕:$\lambda=(A,B,\pi)$。由初始状态概率向量、状态转移概率矩阵 A 和生成概率矩阵 B 决定。π 和 A 决定状态序列,B 决定生成序列。

(4)问题转化。设 S 是所有可能的状态集合,即所有的标签集合 $S=\{S_1,S_1,\cdots,S_k\}$,其中 k 是可能的状态数。设 V 是所有可能的生成属性(用户画像标签)的集合,即读取数据集中每个用户的商品使用情况,判断其属于某属性的可能性。设 O 是生成序列,则 O 的含义为:在标准标签顺序下,用户对已使用过的标签的评分,如 $O=(0.6,0.4,0,0.2)$,0 表示用户没有使用过此标签。如此,*HMM* 模型可在智能推荐问题中转化为预测问题,也称为解码问题,即已知 *HMM* 模型 $\lambda=(A,B,\pi)$ 和 $(O_1,O_2,\cdots,O_n)$ 生成序列,求给定生成序列条件概率 $P(S|O)$ 最大的状态序列 $S=\{S_1,S_2,\cdots,S_k\}$(待推荐标签)。即给定生成序列,求最有可能的对应的状态序列,最终得到最优待推荐的标签序列。

1.1.3 HMM 模型的求解

对 HMM 模型采用维特比算法进行求解。维特比算法实际是运用动态规划求解隐马尔可夫模型预测问题[9],运用动态规划求解概率最大路径(最优路径),而这一条路径对应着一个最佳状态序列。

输入:模型 $\lambda=(A,B,\pi)$ 和生成序列 O;

输出:最优路径 $S=\{S_1,S_2,\cdots,S_k\}$。

(1)初始化:对于每一个可能状态 i,$i=1,2,\cdots,n$,求生成 O_1 的概率:

$$\delta(i)=\pi_i b_i(O_1),i=1,2,\cdots,n \tag{2}$$

(2)递推:对于初始的生成序列,能够计算每一个隐藏状态 i 出现 O_1 的概率,这与初始状态有关,而与转移概率无关。在不同的隐藏状态下,有多条路径可以产生其生成序列。而概率最大的一条路径表示生成此序列 $(O_1,O_2,\cdots,O_n)$ 的概率最大。即:

$$\max_{1\leqslant j\leqslant N}[\delta_{t-1}(j)a_{ji}]b_i(O_t) \tag{3}$$

定义在时刻 t 状态为 i 的所有单个路径 $(i_1,i_2,\cdots,i_{t-1},i_t)$ 中概率最大的路径的第 $t-1$ 个节点为:

$$\psi_t(i)=\arg\max_{1\leqslant j\leqslant N}[\delta_{t-1}a_{ji}] \tag{4}$$

(3)终止:以 P^* 表示最优路径的概率,则:

$$P^*=\max_{1\leqslant j\leqslant N}\delta_T(i),O_T^*=\arg\max_{1\leqslant j\leqslant N}[\delta_T(i)]$$

(4)最优路径回溯,求得最优路径:

$$O^*=(O_1^*,O_2^*,\cdots,O_n^*) \tag{5}$$

1.2 计算矩阵状态转移概率矩阵与生成概率矩阵

同其他大多类型的算法,智能推荐算法的第一步也会对数据进行预处理,而对于不同的数据所采用的处理方式也不尽相同。为提高模型的泛化能力,对于不同的数据情况,本文提出以下几种不同的处理方式,以提高智能推荐系统的泛化能力[10]。因用户存在不同的喜好,而不同喜好导致了选择的标签不同,所以,本文中规定用户喜好为显状态,标签的选择是喜好导致的结果,因此标签设为隐状态。

1.2.1 数据集为 0-1 分布

(1)数据预处理。读取历史信息数据集,计算数据集中不同标签数量的总和,记为 SUM_1,m 为样本数,n 为标签数。为避免用户 U_1 和用户 U_2 的标签记录顺序不同而导致推荐结果错误,将数据集中的标签排列顺序统一到相同标准下:$(x_1,x_2,\cdots,x_n)$,其中 x_i 表示标签属性。如此便得到了规范化的标准标签排序序列。其次将每个用户所含的标签属性数据映射为 0-1 序列(0 表示无,1 表示有),x_{ij} 表示第 i 个样本下第 j 个标签的值,即得到新的数据集。

(2)计算状态转移概率矩阵。在 1.1.2 节中,不难理解状态转移概率矩阵的数学含义为二维离散型随机变量的联合分布律。由于用户的偏好的量化涉及到复杂的自然语言处理和模糊分类,而这对于智能推荐造成了巨大的困难,且用户的偏好会体现到具体的标签属性中,因此,我们使用标签属性和数值来刻画用户属性。基于此,本文提出的状态转移概率矩阵的计算方法如下。首先计算标签 x_j 与标签 x_k 之间的可转移性:

$$x_{jk}=\frac{\sum_{i=1}^{m}(x_j-\bar{x}_j)(x_k-\bar{x}_k)}{\sqrt{\sum_{i=1}^{m}(x_j-\bar{x}_j)^2}\sqrt{\sum_{i=1}^{m}(x_k-\bar{x}_k)^2}} \tag{6}$$

在数据预处理后得到新的数据集，所有数据集中的特征标签$(x_1,x_2,\cdots,x_n)$相互之间进行相关性运算，作为标签之间的相关性度量标准，即为标签 x_j 与标签 x_k 之间的相似性。因推荐相同标签并非刚性需求，因此规定相同标签之间的可转移性为 0，并得到 $n\times n$ 的状态转移概率矩阵。对相似性矩阵中每个元素求和得到可转移总和 SUM，通过式(7)计算得到标签与标签之间对角线为 0 的状态转移概率矩阵 $P_{n\times n}$。

$$P_{ij}=\frac{1}{SUM}\begin{bmatrix}0 & \cdots & x_{1j} & x_{1n}\\ \vdots & \ddots & \vdots & \vdots\\ x_{i1} & \cdots & 0 & x_{in}\\ x_{n1} & \cdots & x_{nj} & 0\end{bmatrix}_{n\times n} \tag{7}$$

且得到的状态转移概率矩阵不同于协同过滤中的数据矩阵，因只在数据中提取标签信息，避免了由样本导致的数据稀疏问题，更多地考虑了标签间的信息，因此能够有效地避免稀疏矩阵的问题。这也是智能推荐准确性和收敛性的保证。

(3)计算生成概率矩阵。1.1.2 节说明了生成概率矩阵的含义为在某种隐含作用下，在某种用户属性下生成为另一标签的概率。根据式(7)，若对状态转移矩阵的每一列进行求和，则求和结果可反映为大数据背景下，由所有标签转移到某一特定标签的概率，即用户对某标签的偏好平均概率分布 $P(X)$。由此，对于标签 j 而言，$P(X_j)$表示标签 x_j 生成某一特定喜好的概率，因难以从标签中精准地推断出用户的喜好，也难以量化为数学形式，借鉴自然语言处理中对偏爱程度的处理形式 Word2Vec[11]，仅以数字表示出对某类偏爱的分布情况，进而能求解出$(x_1,x_2,\cdots,x_n)$标签顺序下的生成概率矩阵 $B=b_j(k)$。

当 $P(X)$值越大，视为对应标签越能普遍反映出用户的偏爱情况，即用户更容易在偏爱的标签消费。而当 $P(X)$值越小时，表示大多数用户普遍厌恶此类标签，此时企业应减少在此类标签的投入成本。

1.2.2　数据集为得分分数

(1)数据预处理。对数据集中的每个样本的标签评分进行标准化处理。Y_j 为第 j 个用户的评分，Y_U 和 Y_D 分别为第 j 个用户对不同标签评分的最高值和最低值。标准化公式为：

$$Y=(Y_j-Y_D)/(Y_U-Y_D) \tag{8}$$

(2)计算生成概率矩阵。对于标准化后的数据，同 1.2.1 节中的方法，将所有样本的标签排列顺序统一到相同标准下：$(x_1,x_2,\cdots,x_n)$。其次，数据集中所有的特征标签相互之间进行求和运算。对于不同用户而言，存在某些用户打分偏低，而有些用户打分普遍偏高的情况。因对数据进行了标准化处理，标准化后的分数的含义仅为某一用户对相应标签的喜好程度，因此避免了不同用户间打分习惯的影响。

事实上，每个标签都有着自身质量属性，而质量属性对用户的相似性计算也会产生一定影响。标签的质量属性是指标签自身质量高、低或者一般。当大多数用户对一个标签的评价一致时，说明该标签并不能区分用户的偏好。当多数用户对一个标签的评价的好坏分布均匀时，即能说明该标签有较好的区分用户偏好的能力。

本文用对标签评分的相异性表示该标签的自身质量属性。标签评分的相异性越高，表明该标签越能区分出用户的偏好，对计算用户相似性的贡献就越大；反之，标签评分的离散程度越低，对计算用户相似性的贡献就越小。考虑到以上情况，以标准差衡量每一个标签的相异度，对于用户的评分数据集，改进的相似性矩阵公式为：

$$\sigma_i=\sqrt{\frac{\sum_{i=1}^{m}(x_i-E(x_i))^2}{m}} \tag{9}$$

$$S=\frac{1}{|\sigma_j+\sigma_k|}\sum_{i=1}^{m}(x_j+x_k) \tag{10}$$

式(9)中，$E(x_i)$为标签 x_i 的评分均值，σ_i 衡量标签 x_i 的相异性。改进的用户相似性计算式(10)考虑了不同标签之间的可信情况，加入了修正因子$|\sigma_j+\sigma_k|$，σ_j 为第 j 个标签的方差，σ_k 为第 k 个样本的方差。若方差越大，则贡献度越低；方差越小，贡献度越高。即考虑了每个服务的贡献度，从而使计算出的用户相似性更加准确。

(3)计算状态概率矩阵。在数据归一化后，计算状态概率矩阵和生成概率矩阵的方法同 1.2.1 节。

1.2.3　数据集为任意数据

(1)数据预处理：不同于 1.2.2 节，评分只为 1～5 分。针对可能存在的特殊情况，如对于标签 L_1，因用户的误操作导致标签 L_1 的方差会偏大，导致计算概率图的结果不准确，对于此，借助箱线图的思想，提出以下数据筛选的解决方案。

箱线图[12]中包括 5 个数：最小值 Min、第一四分位数 Q_1、中位数 M、第三四分位数 Q_3、最大值 Max。在数据集中某些观察值不寻常地大于或小于该数据集中的其他数据，称为疑似异常值。疑似异常值的存在会对后期的计算结果产生不利影响。记第一四分位数 Q_1 与第三四分位数 Q_3 之间的距离为 $IQR(Q_3-Q_1)$，称为四分位数间距。若数据大于 $Q_3+1.5IQR$ 或者小于 $Q_1-1.5IQR$，即认为其是疑似异常值。

(2)在去除疑似异常值之后，同 1.2.1 节中的方法，计算生成概率矩阵和状态转移概率矩阵。

如图 1 所示，针对不同的数据集，提出了不同的处理方式来优化推荐结果，因此模型的泛化能力也得以增强。

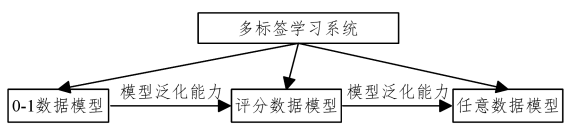

图 1　推荐系统泛化能力示意图

2　智能推荐系统的完善

本文提出了基于 HMM 的智能推荐系统，首先给出算法描述，之后构建用户画像，对算法的结果进行筛选[13]，从而完善智能推荐系统。

用户画像用于勾画用户背景、特征、成员性格、行为场景等，旨在从用户行为数据中挖掘有效信息，尽可能全面细致地抽取一个用户的信息全貌，从而帮助用户从海量数据中享受个性化服务。其是能将定性与定量方式很好地结合在一起的载体，定性通过对用户的生活情景、使用场景、用户心智进行分析来对用户的性质和特征做出抽象与概况；定量可以对特征做精细的统计分析与计算，获得对于用户较为精准的认知，便于用数值刻画用户的特征。

(1)根据建立的新式智能推荐算法对数据集中的标签进行训练，将智能推荐问题转化为标签学习问题，为精确推荐积累学习经验；

(2)当对待推荐的客户进行推荐产品时，同时构建用户画

像，对智能推荐算法推荐的结果进行选择性筛选，得到最终的推荐结果。智能推荐系统示意图如图 2 所示。

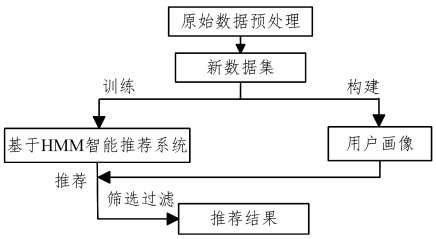

图 2 智能推系统示意图

2.1 用户属性

用户属性也即静态属性，用来描述一个用户的“个性”，从而与其他用户加以区分。因此，为实现精准及个性化的推荐，系统通常对每个用户进行属性建模，其中包括用户的基本信息，如用户的性格、年龄、年收入、活跃时间、所在城市等，而这些静态特征能以定量的形式计算用户静态属性的相似度。

2.2 用户动态属性

用户属性是一种相对静止的数据，缺乏实时性。针对用户属性存在的局限性，我们将建立动态特征[14]。以电影推荐为例，根据用户的观影习惯、观影记录来建立兴趣模型，从而针对用户的爱好进行个性化视频推荐。相比于静态的用户属性，用户动态特征能够更为准确地描述用户特征，是推荐系统中设计用户画像最为重要的信息来源。而这些动态特征能以定性的形式计算用户动态特征的相似度。

用户的一次行为包括人物、时间、地点、事件等基本要素。每一次的用户行为本质上是一次随机事件，可以描述为：什么用户，在什么时间，以什么方式，做了什么行为。“什么用户”涉及对用户的标识；“时间”则包括两个重要的信息，时间戳和时间跨度，其中，时间戳标识用户行为的发生点，时间跨度则标识了用户行为的持续时间；“方式”体现了用户的行为表现，便于根据推荐结果进行推荐；“行为”则直接指示用户的行为偏好，对于精准推荐至关重要。

$$sim=\sum_{n} sta(U_i,U_j) \tag{11}$$

其中，sim 表示以欧氏距离衡量用户和静态属性之间的相似度。

sta 实际操作：对于静态属性的各个标签，年龄、年收入等可量化指标应首先归一化到[0,1]区间内以消除量纲影响。对于城市位置等无法量化的信息，应使用编码措施，对每个城市进行二进制编码，以码距（并规定任意码距不得超过 1，以公平反映静态属性的差异）反映城市位置的差异。对于动态属性，对时间属性、方式属性同样进行归一化处理和编码处理，最终，静态属性相似度的计算方式为：

$$sim=\|\vec{U_i}-\vec{U_j}\|_2=\sqrt{\sum_{u=1}^{n}|U_{iu}-U_{ju}|^2} \tag{12}$$

通过设定阈值的方式[15]便可以对特定的属性进行筛选和约减，筛选出不恰当的推荐结果。静态属性和动态属性共同构成了用户画像，根据用户画像的筛选模型如图 3 所示。

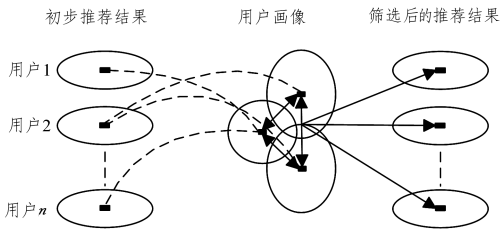

图 3 用户画像工作模型

在构建用户画像后，根据静态属性、动态属性距离进行约减，对每个用户的初步推荐结果进行完善，得到最终的推荐结果。

3 与其他推荐算法的对比分析

本节进行实验验证，将本文所构建的智能推荐方法和传统协同过滤推荐方法进行比较，并分析相关结果。

3.1 实验准备

本文使用的数据集为 Minnesota 大学提供的公开数据，对电影产品的营销推荐进行仿真模拟。为了评估相关算法推荐准确度的差异，选择推荐系统中常用的均方根误差（Root Mean Square Error，RMSE）和覆盖率（Coverage）作为评价标准，计算方法如式（13）和式（14）所示。

$$RMSE=\sqrt{\frac{1}{|T|}\sum_{i=1}^{|T|}(\hat{y}_i-y_i)^2} \tag{13}$$

其中，$|T|$ 表示测试记录次数，i 表示第 i 次测试记录，$\hat{y}_i$ 表示第 i 次记录的预测值，y_i 表示第 i 次记录的真实值。

$$Coverage=g/n \tag{14}$$

其中，n 表示所有的服务数，g 表示用户所有的候选服务中有预测评分的个数。$Coverage$ 越大，预测的效果越好。

3.2 实验结果分析

本文以皮尔逊相关系数为相似性计算方法，基于协同过滤推荐算法（Collaborative Filtering Recommendation，CFR）和本文的多标签学习（Multi Label Learning，MLL）方法进行分析比较。

如图 4、图 5 所示，实验以用户数为横坐标，分别以 RMSE 和 Coverage 评估指标值为纵坐标，观察两种算法的 RMSE 和 Coverage 随用户数量增多的变化情况。注：传统的协同过滤算法处理用户观影时长这样的任意数据有困难，且会导致推荐精度下降，因此这里统一使用 1.2.2 节中的预处理方式对数据进行预处理。

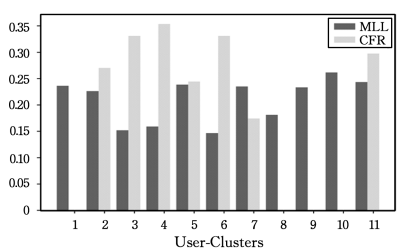

图 4 RMSE 对比

对待推荐用户进行计算，为便于显示将计算结果进行聚类，并对 RMSE 取聚类中心值。由图 4 可见，协同过滤算法由于受到相似度的影响和存在矩阵稀疏的情况，导致对相似度极低的目标用户无法推荐候选服务，造成无法预测的情况，即对无法推荐的用户无 CFR 数据显示；但多标签学习有效地解决了数据稀疏问题，因此能够在保证相似度的前提下进行推荐，证明了 MLL 更适合处理大数据，能更好地在数据稀疏的情况下产生推荐结果，验证了所提方法的可行性。对于某些用户，CFR 和 MLL 同时产生了推荐结果，但 MLL 的结果误差小于 CFR 的误差，证明了 MLL 高效地挖掘了用户信息，保证了推荐结果的准确性。

在图 5 中，两种算法的覆盖率随着用户数量的增加逐渐增高。其中 MLL 方法的覆盖率远远超过了 CFR 方法，更能有效地对用户喜好进行推荐。同时，因为计算了状态转移概率矩阵，能够很好地根据用户历史记录对用户推荐新的标签，

由此可见，本文提出的多标签学习的智能推荐方法可以产生更好的推荐效果。

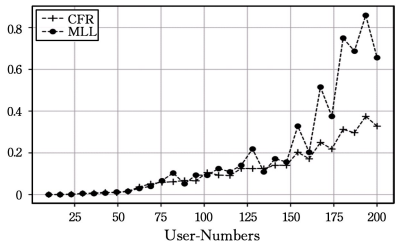

图 5　Coverage 实验结果的对比

4　算法收敛性检验

一个优良的算法应该具有稳定的收敛性，如神经网络、遗传算法等都要考虑结果的收敛性以保证结果的正确。而本文的多标签推荐过程并不存在迭代计算的情况，因此为表示 MLL 推荐中的迭代过程，以用户数为迭代变量，即检验随用户数量增加算法是否收敛。同样以 RMSE 为误差衡量标准，通过增加用户数量观察 RMSE 的变化。

与此同时，因传统的推荐算法存在数据稀疏的问题，导致推荐结果不够稳定，所以在用户数量迭代的过程中对 MLL 推荐算法的 RMSE 进行检验，结果如图 6 所示。

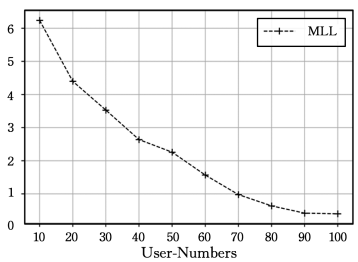

图 6　算法收敛性检验

如图 6 所示，随着用户数量增加至 100 人左右，多标签计算避免了数据稀疏问题，并且设置标签为隐状态，用户偏好为显状态。首先通过生成概率矩阵度量两状态的联系，然后通过用户画像更加精细联系，因此多标签推荐算法的推荐误差逐渐减小。而传统的推荐算法因数据稀疏和只计算一次标签相似性，导致推荐结果产生了震荡而不够稳定。

最终经过多次验证，在用户数量超过某一阈值时，传统推荐算法和本文的多标签推荐算法的误差开始收敛，此时认为推荐结果可以接受。但在小于该阈值时，本文的多标签推荐的结果更加稳定，即算法的收敛性得到了验证。

结束语　本文首先提出了一种基于多标签学习的智能推荐系统，针对不同类型的原始数据进行预处理，将智能推荐与标签学习进行有机结合，挖掘处理后的数据中标签间潜在的联系，获得多标签学习的经验。采用概率图的方式考虑了相似用户属性，同时权衡了自身属性的因素，基于概率推理机制，度量用户或标签之间的相似性；之后，运用隐马尔可夫模型学习多标签学习得到的经验，求得最优的待推荐标签顺序，即可获得推荐的商品服务顺序，并且有效地解决了数据稀疏问题。最终，根据用户信息构建用户画像对推荐特征进行有效的筛选，从而提供个性化服务，进一步满足用户的需求，提高了推荐服务的质量。

本文所构建的智能推荐系统自成一体，从数据预处理到智能推荐算法中引入多标签学习的概念，以概率图的形式进行学习和训练，避免了传统协同过滤算法的众多弊端，最终以用户画像对推荐结果进行修正，使之可以有效地实现智能推荐的目的，成为完备智能推荐系统。由于引入了概率图的理论，相对于传统的协同过滤，其在处理相同的数据量时，保证了算法收敛性，同时只需要消耗较少的计算资源。

参 考 文 献

[1] FRANCESCO R，LIOR R，BRACHA S. Recommender Systems Handbook[M]. Boston，MA：Springer US，2011.

[2] 吴琛. 基于 K-means 聚类算法的协同过滤服务质量预测[D]. 杭州：浙江大学，2017.

[3] YAO Y，TONG H H，YAN G. Dual-Regularized One-Class Collaborative Filtering[C]// Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management. New York：ACM，2014：759-768.

[4] ZHANG M，ZHOU Z. A Review on Multi-Label Learning Algorithms[J]. IEEE Transactions on Knowledge and Data Engineering，2014，26(8)：1819-1837.

[5] YANG D X，LIU Z J，ZHOU J L. Chaos optimization algorithms based on chaotic maps with different probability distribution and search speed for global optimization[J]. Communications in Nonlinear Science and Numerical Simulation，2014，19(4)：1229-1246.

[6] CHAI M L，WANG L D，WENG Y L，et al. Single-view hair modeling for portrait manipulation[J]. ACM Transiation Graph，2012，31(4)：116.

[7] HU J，YU X，QIU D，et al. A simple and efficient hidden Markov model scheme for host-based anomaly intrusion detection[J]. IEEE Network，2009，23(1)：42-47.

[8] LEE Y S，CHO S B. Activity Recognition Using Hierarchical Hidden Markov Models on a Smartphone with 3D Accelerometer[M]// Hybrid Artificial Intelligent Systems. Berlin：Springer，2011.

[9] DJUROVIĆ I. Viterbi algorithm for chirp-rate and instantaneous frequency estimation[J]. Signal Processing，2011，91(5)，1308-1314.

[10] WANG X Z，SHAO Q Y，MIAO Q. Architecture selection for networks trained with extreme learning machine using localized generalization error model[J]. Neurocomputing，2014，102(1)：3-9.

[11] ZHANG D W，XU H，SU Z C. Chinese comments sentiment classification based on word2vec and SVMperf[J]. Expert Systems with Applications，2015，42(4)：1857-1863.

[12] ANDROIC D，ASATURYAN D S，ASATURYAN A，et al. First Determination of the Weak Charge of the Proton[J]. Nature，2018，557：207-211.

[13] 蔡海尼，陈程，文俊浩，等. 基于用户签到和地理属性的个性化位置推荐算法研究[J]. 计算机科学，2016，43(12)：163-167，178.

[14] SINGH R，MILLER S A，ROWLANDS A S. Dynamic characteristics of a direct-heated supercritical carbon-dioxide Brayton cycle in a solar thermal power plant[J]. Energy，2013，50(1)：194-204.

[15] ORIOL V，CHARLES B，TIMOTHY L. Matching Networks for One Shot Learning[J]. Advances in Neural Information Processing Systems，2016，29(1)：3630-3638.