# 数据挖掘中的推荐算法综述

耿鑫[1],刘晋佩[2]

(1.同济大学 计算机科学与技术系,上海 201804;2.厦门大学 信息科学与技术学院 自动化系,福建 厦门 363105)

**摘要:**推荐算法是推荐系统的核心,近年来,推荐系统受到了研究人员和学术界的关注,到目前,研究人员提出了很多推荐算法。该文侧重讨论现有的推荐算法及其性能,并在此基础上,进一步提出了未来的研究方向。

**关键词:**数据挖掘;推荐系统;协同过滤

**中图分类号:TP 301 文献标识码:A 文章编号:1009-3044(2012)19-4691-06**

**Survey of Recommender Algorithms in Data Mining**

GENG Xin[1], LIU Jin-pei[2]

(1.Department of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2.Department of Automation, Xiamen University, Xiamen 363105, China)

**Abstract:** The recommender algorithm is the major part of the recommender system. In recent years, researchers and academics have paid more attention to the recommender system and proposed a lot of recommendation algorithm. This paper focuses on discussion of the existing recommendation algorithm and its performance, and on this basis, the future research directions.

**Key words:** data mining; recommender system; collaborative filtering

## 1 概述

在知识爆炸的互联网时代,网络的信息量以指数级的速度在不断增长,简单搜索引擎已无法满足用户在海量信量中获取信息的需要,信息的利用率降低。为解决这一问题,研究人员提出了推荐系统,可有效地解决在海量信息中的获取有用信息的问题。推荐系统是通过分析用户的历史行为,提示用户习惯和喜好,建立相应的推荐算法,为每个用户产生一个推荐列表,使其可以快速地找到自己感兴趣的信息。

上世纪末,推荐系统主要应用于音乐、电影、书籍等产品的推荐。近年来,推荐系统已被广泛地应用于电子商务领域,成为电子商务中不可缺少的一部分,各大电子商务网站,如 Amazon、taobao、ebay 都不同程度的使用了推荐系统,显著地提高了电子商务企业的销售额,同时也为用户搜索商品提供了方便。与此同时对推荐系统的研究在理论上促进了多学科的交叉发展。设计出更优秀的推荐算法已经成为理论界关注的热点。目前为止,学者们提出了基于内容的推荐算法、协同过滤推荐算法、基于复杂网络的推荐算法、混合推荐算法等不同的算法,数据挖掘领域、机器学习领域的一些新方法也被应用到推荐算法中。

## 2 推荐系统及其关键问题

文献[1]给出推荐系统非形式化定义:"利用电子商务网站向客户提供商品信息和建议,帮助客户决定购买的产品,模拟销售人员帮助客户完成购买。"图1给出了推荐系统模型流程,主要过程是收集用户喜好数据和对用户产生推荐两个部分,收集用户喜好数据的方式有两种,一种是显性收集,即向用户主动的收集,例如要求用户填写调查表格;另一种是隐性收集,即收集记录用户行为的历史数据,如浏览过的网页、搜索过的关键词、购买过的物品。对用户产生推荐是推荐系统中最核心的部分,使用推荐算法或推荐模型对用户喜好信息及产品信息进行计算和处理,每个用户均会得到一个产品推荐列表,取推荐列表中排名靠前的产品得到推荐结果,并将推荐结果返回给用户。

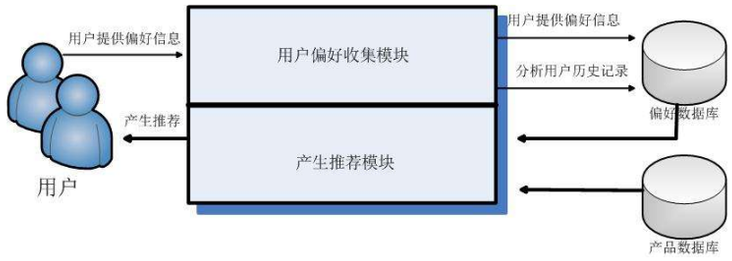


**图1 推荐系统模型**

收稿日期:2012-05-22
基金项目:国家自然科学基金(60972036)

### 2.1 推荐系统评价

1)预测准确度评价指标

很多推荐系统都用准确度来评价推荐算法的好坏,准确度为推荐算法预测排名与实际排名之间的差异度。预测准确度一个常用的方法是度量推荐系统预测打分与实际打分的平均绝对误差(Mean Absolut Error, MAE)[2,20]:

$$MAE=\frac{1}{c}\sum_{\alpha=1}^{c}\left|v_{i\alpha}-r_{i\alpha}\right|$$

其中,c为系统中被用户i打过分的产品个数,$r_{i\alpha}$为用户对产品的实际打分,$v_{i\alpha}$为推荐系统的预测打分。平均绝对误差的优点在于计算方法简单及可直观比较两个系统预测准确度。

其他类似的评价指标还有均方根误差(Root Mean Squared Error, RMSE)[3]:

$$RMSE=\sqrt{\frac{1}{n_t}\sum_{(i,\alpha)}\left|v_{i\alpha}-r_{i\alpha}\right|^2}$$

其中,$n_t$为系统中用户-产品对($i$, $\alpha$)的个数,

2)分类准确度评价指标

分类准确度[4]是衡量推荐算法中用户的喜好判定正确的比例。分类准确度指标有准确率和召回率,准确率表示用户对被推荐产品感兴趣的概率,准确率的计算方法为:

$$P=\frac{N_{rs}}{N_s}$$

其中,$N_s$为推荐给用户的总产品数,$N_{rs}$为推荐给用户的产品中,用户喜欢的产品数。

召回率表示一个用户喜欢的产品被推荐的概率,其计算方法为:

$$R=\frac{N_{rs}}{N_r}$$

其中,$N_r$为系统中总的产品中用户喜欢的产品数。

另一个度量系统分类准确度的重要指标是ROC曲线[5,6]。

3)排序准确度

排序准确度是用来度量推荐结果列表与用户对产品排序列表之间的吻合度。周涛[7]等提出了排序准确度具体定义:

$$r_i=\frac{L_i}{N}$$

为了衡量推荐算法的好坏,推荐系统将评分数据集的90%划分为训练集,10%划分为测试集,公式中,N为训练集中用户没有选择过的产品个数,$L_i$为测试集中待测的产品i在推荐列表中的排名,因此$r_i$越小,说明是将用户喜欢的产品排到了前面,即排序准确度越高。

## 3 推荐算法

### 3.1 基于内容的推荐算法

基于内容的推荐算法是根据用户喜好及习惯的记录和用户已经选择过的产品内容信息来计算用户的信息与用户没有选择过的产品信息的相似度,然后对用户产生推荐。例如,在音乐推荐系统中,为向用户推荐其可能感兴趣的音乐,系统需要收集该用户曾经听过的音乐的共同特点,如说演唱者、音乐类型、歌曲语言及歌曲年代等,基于这些共性,向用户推荐与用户喜欢的音乐相近的音乐。

基于内容的推荐本质上属于信息获取和信息过滤,以往的研究主要关注了基于文本的信息过滤方面,因此当前基于内容的推荐系统都是通过分析特定产品的文本信息来进行推荐的。

现在许多基于内容的推荐系统在文本信息过滤的基础上进行了改进,通过建立用户的偏好配置文件,来记录用户的喜好信息。

用户的偏好配置文件通常通过显性和隐性两种方式获得,显性方式指用户需要填写调查问卷来确定用户的兴趣偏好,隐性方式指用户不需要专门来回答问题,由系统自动收集用户浏览或购买习惯信息。

在获得用户偏好配置文件及待推荐产品信息文件的过程中,通常采用TF-IDF( term frequency inverse document frequency)[8]方法来表征文件。例如FAB网页推荐系统中,页面收集代理从web上收集特定主题的页面,个人推荐代理从特定主题中选择出用户感兴趣的页面推荐给用户[3]。

基于内容的推荐系统中除使用信息获取的技术,还有其他的一些技术,典型的有贝叶斯分类以及决策树[9,10]。与基于信息获取方法不同的是,这些算法并不是基于一个函数公式,而是从统计学角度出发或利用机器学习技术从已有的数据中通过分析得出模型,然后基于此模型产生推荐。例如,用户将一组网页按照自己的爱好分为感兴趣的和不感兴趣的,利用贝叶斯分类器就可以分辨出用户对此网页是否感兴趣,贝叶斯分类器可处理大型的数据,且方法简单、分类准确率高、速度快。具体来说就是贝叶斯分类器可用来估计一个网页P属于某个类C(用户感兴趣或不感兴趣的类)的概率。实验结果表明这种分类方法在实际应用中有较高的分类准确率[10]。

由于基于内容的推荐系统中最为核心的部分是用户的偏好配置文件的构建和更新。Somlo和Howe[11]等人把自适应过滤技术应用到更新用户偏好配置文件中,基本思路是通过对用户兴趣的信息获取构建用户偏好配置文件,然后把产品信息文件与用户偏好配置文件进行对比,把与用户偏好配置文件相似度高的产品信息文件推荐给用户的同时更新用户偏好配置文件。

在此基础上,Robertson和Walker等人[12]在自适应过滤的基础上设定一个最低阈值,其基本思想是相似度大于一定值的产品信

息文件才可以用于更新用户偏好配置文件。通过这种方法,不仅有效的提高了推荐算法的准确性,而且大幅度地提高了系统的运行效率。

文件中的关键字的同义和多义现象会导致相似度的计算不准确。对于这个问题,文献[13]提出了潜在语义分析方法(Latent Semantic Analysis, LAS),采用文档-词矩阵奇异值分解的方法将文档和词映射到同一个低维的潜在语义空间中,用户提出的查询也会被映射到这个语义空间,在这个空间可以更准确地计算偏好配置文件与产品信息文件的相似度。

但LSA采用奇异值分解得到的低维语义空间物理意义不明确,同时矩阵的奇异值分解计算量大。文献[14]提出了概率潜在语义分析( probabilistic latent semantic analysis, PLSA)模型较好地克服LSA模型的不足。文献[14-15]中的实验表明,基于PLSA模型的推荐算法准确度优于聚类模型、贝叶斯模型和皮尔森相关性的推荐算法。

### 3.2 协同过滤推荐算法

协同过滤算法是目前应用最广泛最成功的推荐算法[16],协同过滤算法基于这样的假设:用户对一些相似产品的评分比较相似;兴趣爱好相似的用户群对一些产品的评分比较相似。因此协同过滤算法的基本思想是根据用户对产品的历史评价来计算用户间的相似度,将与目标用户相似度高的用户组成目标用户的邻居集,根据邻居集对产品的评价来对目标用户产生推荐。

目前已有很多协同过滤推荐系统,第一个真正投入到实际应用的协同过滤推荐系统是Grundy[17],该系统通过对用户兴趣建模,利用模型向用户推荐书籍。 随后又出现了Tapestry,GroupLens以及Ringo等协同过滤推荐系统。协同过滤推荐算法可分为两类:基于记忆的推荐算法与基于模型的推荐算法。

1)基于记忆的协同过滤算法

基于记忆的协同过滤算法是根据已经评价过的产品信息进行预测。假设用户集为$C=\{c_1,c_2,\cdots,c_n\}$,产品集为$S=\{s_1,s_2,\cdots s_m\}$。设$r_{c,s}$为用户c对产品s的预测评分,$r_{c,s}$是根据其他用户对s的评价计算得出的。设C*为用户c的邻居集,那么可用如下算法预测$r_{c,s}$的值[18]:

$$r_{c,s}=\frac{1}{n}\sum_{\bar{c}\in C^*}r_{\bar{c},s} \tag{1}$$

$$r_{c,s}=k\sum_{\bar{c}\in C^*}sim(c,\bar{c})\times r_{\bar{c},s} \tag{2}$$

$$r_{c,s}=\overline{r_c}+k\sum_{\bar{c}\in C^*}sim(c,\bar{c})\times(r_{\bar{c},s}-\overline{r_{\bar{c}}}) \tag{3}$$

其中k是标准化因子,一般$k=1\Big/\sum_{\bar{c}\in c^*}|sim(c,\bar{c})|$,sim(i,j)表示用户i和用户j的相似性。$\overline{r_c}$表示用户c已经评价过的产品的平均评分。公式(1)是最简单的预测算法,即直接计算用户邻居集对产品评分的平均值。公式(2)是最常用的加权平均算法,不只考虑邻居的评分,还考虑邻居对目标用户的影响力,相似度高的邻居对目标用户的影响力就大。由于(2)中加入了标准化因子k,所以可用于计算不同的推荐系统中的用户相似度。但公式(2)并没有考虑不同用户评价习惯的不同,例如有些用户评价普遍偏高,有些用户评价普遍偏低。因此公式(3)考虑了不同用户平均喜好程度的念头,解决了评价尺度不同的问题,预测的准确度也较公式(2)高。

协同过滤推荐算法中十分重要的一个步骤就是邻居集的确定,通过用户对产品的评分信息计算用户间的相似度来确定目标用户的邻居集。计算相似度最常用的方法是Pearson相似性[19]和余弦相似性[21],它们均定义用户x和y共同打过分的产品集合为:$s_{xy}=\left\{s\in S|r_{x,s}\neq\varphi r_{y,s}\neq\varphi\right\}$。基于Pearson相关性的x和y相似度定义为:

$$sim(x,y)=\frac{\sum_{s\in S_{xy}}(r_{x,s}-\overline{r_x})(r_{y,s}-\overline{r_y})}{\sqrt{\sum_{s\in S_{xy}}(r_{x,s}-\overline{r_x})^2\sum_{s\in S_{xy}}(r_{y,s}-\overline{r_y})^2}}$$

基于余弦相似性算法中,将用户x和y对产品的评分信息用m维向量表示,则二者的余弦值为:

$$sim(x,y)=\cos(\vec{x},\vec{y})=\frac{\vec{x}\cdot\vec{y}}{\|\vec{x}\|_2\times\|\vec{y}\|_2}=\frac{\sum_{s\in S_{xy}}r_{x,s}r_{y,s}}{\sqrt{\sum_{s\in S_{xy}}r_{x,s}^{\ 2}\sum_{s\in S_{xy}}r_{y,s}^{\ 2}}}$$

$\vec{x}\cdot\vec{y}$表示向量的点积,$\|\vec{x}\|_2$表示向量的模。为了提高响应时间,所有用户的相似性并不是实时计算,而是提前计算并保存在数据库中,以一定的时间间隔进行更新。

目前还有很多改进的算法应用到相似性计算中,如默认投票(default voting)、事件扩展(case amplification) [20]、主要权重预测(weighted-majority predition)等[21]。其中,缺席投票是对基于记忆方法的扩展,缓解了数据稀疏问题,其主要思想是给用户没有评分或评分很少的产品赋于缺省的评分值。该算法大幅度提高了相似性计算和预测评分的准确性。Sarwar等人[22]提出了基于产品相似性的方法,利用产品相似性进行评分。Deshpande和Karypis[23]在基于产品相似性的基础上提出了top-N推荐算法,即只考虑相似度最高的N个产品。实验结果表明该推荐算法不仅远小于基于用户推荐算法的计算复杂度,且因为排除了不太相似产品的影响,提高了算法推荐准确性。Yang和Gu[24]给用户的短期兴趣赋予更高的权重,实验证明此方法的推荐准确性比传统基于用户相似性的推荐算法高。

2)基于模型的协同过滤算法

基于模型的协同过滤算法基本思想是对已有的用户评分数据应用统计学和机器学习得到模型,通过模型对用户产生推荐。Breese等[20]人提出基于概率模型的协同过滤算法,其预测评分可形式化的描述为:

$$r_{c,s}=E(r_{c,s})=\sum_{i=0}^{n}i\times\Pr(r_{c,s}=i|r_{c,\bar{s}}\in S_c)$$

预测评分值为0到n之间的整数。概率Pr表示根据用户过去对产品的评分，用户对产品s的预测评分为由此公式计算出的评分的概率。Breese等人使用聚类模型和贝叶斯网络两个模型来计算概率。

一些其他推荐算法模型和方法有统计模型、概率相关模型、线性回归模型、最大熵模型、机器学习方法、马尔可夫决策过程方法以及将推荐看作序列决策问题等。

### 3.3 基于二部图关系的推荐算法

二部图是复杂网络中一个概念，对于复杂网络，若网络中的节点是同一类节点，则这种网络称为单模式网络，而有些网络由两种类型的结点构成，同一类的结点不能直接相连，只有不同类的结点可以连接，这种网络称为二部图网络。

在推荐系统中，用户与产品形成的网络即可看作二部图网络。周涛等人[25]提出了基于二部图网络的资源分配方案。假设二部图中每个结点都有一定的“资源”，推荐系统中这种资源可以理解为推荐的影响力。资源分配过程如下：

设二部图网络G(U,I,E)，E为边的集合，用户U和产品I中的结点分别为$U_1, U_2, \cdots, U_n$和$I_1, I_2, \cdots, I_m$。U中第i个结点的初始资源为$f(U_i) \geq 0$。所有的资源都等概率地从U流向I，那么位于I中第l个结点的资源为：

$$f(I_l) = \sum_{i=1}^{n} \frac{a_{il} f(U_i)}{k(U_i)}$$

其中$k(U_i)$为结点$U_i$的度，$a_{il}$为$n \times m$邻接矩阵：

$$a_{il} = \begin{cases} 1, U_i I_l \in E, \\ 0, U_i I_l \notin E, \end{cases}$$

然后，流向I的资源再回流到U，此时结点$U_i$的资源为：

$$f'(U_i) = \sum_{j=1}^{n} w_{ij} f(U_j)$$

其中

$$w_{ij} = \frac{1}{k(U_j)} \sum_{l=1}^{m} \frac{a_{il} a_{jl}}{k(I_l)} \tag{4}$$

$w_{ij}$可被理解为$U_j$愿意分配给$U_i$的资源。整个过程可表示为$\vec{f'} = W\vec{f}$，最终将向量$\vec{f'}$中用户没有选择过的数据值进行降序排序，数值大表示用户喜欢的概率大，因此排靠前得会被推荐给用户。

另一种计算矩阵W的方法是：

$$w_{ij} = \frac{1}{k(U_i)} \sum_{l=1}^{m} \frac{a_{il} a_{jl}}{k(I_l)} \tag{5}$$

式(5)更倾向于向用户推荐多样的、不流行的产品，而(4)式得到的推荐列表倾向于向用户推荐流行产品。

在基于二部图推荐系统中，越流行的产品其推荐能力超强，这在一定程度上会降低推荐的多样性和准确性。周涛等人通过抑制流行产品的推荐能力来提高推荐的准确性，即在产品初使资源表达式中加入自由参数$\beta$，利用$\beta$的值来控制流行产品的推荐能力，对于任意目标用户$U_i$，设初资源为：

$$f(I_j) = a_{ji} k_j^{\beta}$$

当$\beta > 0$时，度大产品的推荐能力加强，当$\beta < 0$，度大产品的推荐能力被消弱。实验表明当$\beta = -0.8$时，推荐准确性最高。

另一种提高推荐准确性的方法是去除重复属性，周涛等人通过考虑用二阶耦合，从关联矩阵中适当得减去二阶关联，重新定义关联矩阵：

$$W' = W + \alpha W^2$$

资源分配为$\vec{f'} = W'\vec{f}$，$\alpha$为可调参数。实验表明$\alpha = -0.75$时，算法准确性最高。

由于目标用户与大多数用户之间选择同一产品的概率较小，因此目标用户只与一小部分其他用户的关联性较强，为了提高算法的准确性并降低算法的时空复杂度，可以引入耦合阈值来降低运算的复杂度。文献[26]提出在建立二部图前先计算用户的相似性，用相似性大于阈值的用户组成二部图网络，这样得到规模较小的子二部图网络，计算量得到了降低。实验表明，此方法降低了推荐算法的运算时间复杂性，同时提高了推荐算法的准确性。

### 3.4 混合推荐

由于上述各大算法均存在各自的不足，因此出现了混合推荐算法，通过不同的组合方式把已有的推荐算法组合在一个模型中，达到优势互补的目的。主要的混合推荐算法有以下几种形式：

1)组合不同推荐算法的推荐结果

此方法基本思想为分别用基于协同过滤推荐算法和基于内容的推荐算法对目标用户产生推荐结果。然后将结果进行线性组合，产生综合推荐结果，推荐给用户[27]。还有一种方法是按照一个或几个推荐算法评价的指标，选择相比较好的推荐算法来产生推荐，比如DailyLearner系统[4]。

2)基于内容的算法加入基于协同过滤的算法

这种方法是利用降维技术把基于内容的对象特征进行简化。Soboroff等人[28]利用潜在语义索引(LSI)算法，在基于内容框架中使用精华的特征向量来产生推荐。得到了比只使用基于内容的算法更好的推荐结果。

3)基于协同过滤的推荐算法加入基于内容的推荐算法

Fab推荐系统[29]就使用了这种推荐算法，总体过程是按照协同过滤的算法进行，但是在计算邻居集时，用户相似度是根据用户

的配置文件,利用基于内容的推荐算法得到的。显著减弱了协同过滤算法中数据稀疏性的影响。

4)基于内容和协同过滤混合算法

Schein等[30]利用概率浅层语义分析提出了结合基于内容和协同过滤的统一概率推荐方法。其思想是将用户感兴趣的信息表示成主题,利用全概率公式对用户感兴趣的主题进行预测。Condiff等[31]提出Bayes混合效用模型对未知产品进行预测。Ansari等人[32]利用用户和产品的配置文件,利用统计模型来预测用户i对产品j的评分:

$r_{ij}=x_{ij}\mu+z_i\gamma_j+\omega_j\lambda_i+e_{ij}$,其中 $e_{ij}\sim N(0,\sigma^2)$ , $\lambda_i\sim N(0,\Lambda)$ , $\gamma_j\sim N(0,\Gamma)$

$e_{ij}$ , $\lambda_i$ 与 $\gamma_j$ 是分别表示噪声效应,用户的异质性及产品的异质性的随机变量。3种分布的参数 $\sigma^2$ , $\Lambda$ , $\Gamma$ 是由马尔可夫蒙特卡洛方法估算得到的。

## 4 推荐算法的性能比较

推荐算法由于算法思想的不同,都各自有各自的优点与不足,例如基于内容的推荐算法根据用户的偏好配置文件与产品信息文件来计算相信性,不需要其他用户的数据,因此不存在冷启动问题,还可以发现用户特有的兴趣,推荐不流行的产品。用户的偏好配置文件或产品的信息文件可以用来描述他们各自的特征,用于计算相性并产生推荐,因此基于内容的推荐不存在数据稀疏问题,但目前技术对文本文件信息提取较为容易,对图像或音频的数据提取技术还不是很成熟,因此基于内容的推荐算法会受产品类型的限制。而协同过滤算法不受产品类型的约束,可推荐视频流或音频流产品,协同过滤推荐算法依赖用户历史数据,因此对老用户的推荐准确度较高,并善于发现用户的新兴趣。但由于十分依赖历史数据,对于新用户或新产品无法产生推荐,同时系统用户和产品都不在断的增长,历史数据会显得越来越稀疏,系统的整体性能也随之下降。表1具体列出了几种推荐算法的优点与不足。

**表1 推荐算法对比表**

| 推荐算法 | 优点 | 不足 |
|---|---|---|
| 基于内容的推荐算法 | 不存在冷启动问题;<br>不存在数据稀疏性问题;<br>可发现用户特有的兴趣,推荐不流行的产品; | 只可推荐配置文件为文本信息的产品;<br>难以发现用户的新兴趣; |
| 协同过滤推荐算法 | 不受产品类型的约束,可推荐多媒体产品;<br>对老用户推荐准确度高;<br>可发现用户的新兴趣; | 存在冷启动问题;<br>可扩展性差;<br>存在数据稀疏性问题; |
| 基于二部图推荐算法 | 开辟了推荐算法新的研究方向; | 存在冷启动问题;<br>存在数据稀疏性问题; |

## 5 推荐算法开放性课题及进一步研究

### 5.1 数据稀疏性问题

依赖用户对产品的评价信息的推荐系统中,数据稀疏是推荐系统面临的主要问题。评分矩阵的稀疏会严重影响推荐算法的性能。例如在协同过滤推荐算法中,刚加入系统的用户或产品,在新用户或产品评价信息较少的情况下,计算其邻居集是非常困难的。

降维技术[4,33]是解决稀疏性问题常用的一种方法。采用奇异值分解方法删除次要的用户和产品,达到降低矩阵维数的目的。文献[34]中通过整合产品间的关系,将用户-产品矩阵转换成用户-类别矩阵,有效得降低评分矩阵的稀疏度。还有人采用潜在语义索引技术[13-15]将用户或产品投影到低维的空间上,再计算相似度,在一定程度上缓解了数据稀疏问题。

### 5.2 可扩展性问题

由于推荐系统用户数和产品数的增多,要找到适合某个用户感兴趣的商品的难度也就越大,可扩展问题成为推荐算法面临的严重问题。采用降维、聚类和分类等策略可在一定程度上解决可扩展性问题。SVD等降维技术可压缩矩阵,产生较好的推荐结果,但分解矩阵过程会消耗大量时间;基于最近邻KNN算法[35],只考虑相似度高于某一阈值的邻居集,在一定程序减少了时间的开销。

### 5.3 其他问题

除以上问题外,推荐系统面临的其他问题。例如,较难提取多媒体数据(图形、视频流、无声音流等)的部分特征。还有隐私问题,很多用户不愿意对产品进行评价,因害怕信息泄漏而不愿意提交更多个人信息。

## 6 结论

由于互连网的迅猛发展,信息过载已经是一个不可忽略的问题,因此对推荐系统的需求也与日俱增。经过这些年的发展,推荐系统在研究和应用中已有了长足的进步,但仍然有数据稀疏、可扩展、冷启动、特征提取难等问题。通过对推荐算法的分析可以看到每一种算法都有优势和弊端,没有单独一种算法或改进算法很好的解决目前推荐系统存在的问题,因此,混合推荐算法会成为研究的一个趋势,根据不同的需求来组合推荐算法,使其对于需求具有最好的性能。伴随着推荐算法的研究和发展,推荐系统会帮助用户更便捷、有效地获取有用信息。

## 参考文献:

[1] Resnick P,Varian H R.Recommender Systems[J]. Communications of the ACM,1997,40(3):56-58.

[2] Shardanand U,Maes P.Social Information Filtering Algorithms for Automating "word of mouth" [C]//Proc. of ACM CH I' 95 Conference on Human Factors in Computing Systems. New York ACM Press,1995:210-217.
[3] Balabanovic M,shoham Y.Fab: Content-based Collaborative Recommendation[J].Communications of the ACM,1997,40(3):66-72.
[4] Billsus D,Pazzani M J.Learning Collaborative Information Filters[C]//Rich C.Mostow J.Proc. of the 15th National Conference on Artificial Intelligence (AAAI-1998). Menlo Park, Calif AAAI Press,1998:46-53.
[5] Swets J A.Information Retrieval Systems[J].Science,1963,141:245-250.
[6] Swets J A.Effectiveness of Information Retrieval Methods[J].Amer Doc,1969(20):72-89.
[7] Zhou T,Jiang L L,Su R Q,et al.Effect of Initial Configuration on Network-based Recommendation[J].Europhys Lett,2008,81:58004.
[8] Salton G.Automatic Text Processing [M]. New York:Addison Wesley ,1989.
[9] Park H S,Yoo J O,Cho S B.A Context aware Music Recommendation System Using Fuzzy Bayesian Networks with Utility Theory[J].Fuzzy System and Knowledge Discovery ,2006,4223:970-979.
[10] Mooney R J,Bennett P N,Roy L.Book Recommending Using text Categorization with Extracted Information [C]//Proc. Recommender Systems Papers from 1998 Workshop,1998:70-74.
[11] Somlo G,Howe A.Adaptive Lightweight Text Filtering[C]//Proc.Lecture Notes in Computer Science,2001:319-329.
[12] Robertson S,Walker S.Threshold Setting in Adaptive Filtering[J]. Documentation, 2000(56):312-331.
[13] Deerwester S,Dumais S,Furnas G,et al.Indexing by Latent Semantic Analysis[J].The Journal of the American Society for Information Science,1990,40( 6):391-407.
[14] Hofmann T.Latent Semantic Models for Collaborative Filtering[J].ACM Transaction Information Systems,2004,22(1):89-115.
[15] Hofmann T.Unsupervised Learning by Probalilistic Latent Semantic Analysis[J].Machine Learning, 2001,42:177-196.
[16] Jonathan L,Herlocaker,Joseph A.Konstan,John Riedl.Explain in Collaborative Filtering Recommendations[C]//Proc. ACM conference on Computer supported cooperative work,2000, 241-250.
[17] W. Hill, L. Stead, M. Rosenstein, G Furnas. Recommending and Evaluating Choices in a Virtual Community of Use[C].//Proc. of Conference Human Factors in Computing Systems, 1995, 194-201.
[18] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions[J]. IEEE Trans. on Knowledge and Data Engineering, 2005, 17(6): 734-749.
[19] Tversky, Amos, Features of Similarity[J], Psychological Review. 1977, 84(4): 327-352.
[20] Breese J, Hecherman D, Kadie C. Empirical Analysis of Predictive Algorithms for Collaborative Filtering[C].//Proc. of the 14th Conference on Uncertainty in Artificial Intelligence(UAI'98), 1998, 43-52.
[21] Delgado J, lshii N. Memory-Based Weighted-majority Prediction for Recommender Systems.//Proc. of the ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation. New York: ACM Press. 1999.
[22] Sarwar B, Karypis G Konstan J, et al, Item-BaSed Collaborative Filtering Recommendation Algorithms[C].//Proc. 10th Int'l WWW Conf., Hong Kong, 2001, 1-5.
[23] Deshpande M, Karypis G Item-Based Top-N Recommendation Algorithms[J]. ACM Trans. Information Systems, 2004, 22(1): 143-177.
[24] Yang MH, Gu ZM, Personalized Recommendation Based on Partial Similarity of Interests[C].//Proc. Advanced data mining and applications, 2006(4093), 509-516.
[25] Zhou T, Ren J, Medo M, Zhang Y C. Bipartite Network Projection and Personal Recommendation[J]. Phys. Rev. Lett, 99(2007) 154301.
[26] Kuscsik Z, Zhang Y C, Zhou T, Improved Recommendation Algorithm with Similarity Threshold[J]. submitted to Phys. Rev. E.
[27] Claypool M, Gokhale A, Miranda T, et al, Combining Content-Based and Collaborative Filters in Online Newspaper[C].//Proc. ACM SIGIR'99 Workshop Recommender Systems: Algodthms and Evaluation, Berkeley, 1999, 366-375.
[28] Soboroff I, Nicholas C. Combining Content and Collaboration in Text Filtering[C].// Proc. of the Int'l Joint Conf. on Artificial Intelligence Workshop: Machine Learning for Information Filtering. Stockholm, 1999. 86-91.
[29] Pazzani M, Billsus D. Learning and Revising User Profiles: The Identification of Interesting Web Sites[J]. Machine Learning, 1997(27): 313-331.
[30] A. I. Schein, A. Popescul, L. H. Ungar, and D.M. Permock. Methods and Metrics for Cold.-Start Recommendations[C].//Proc. of the 25th annual international ACM SIGIR conference on Research and development in information retrieval, 2002, 253-260.
[31] Condliff M,Lewis D,Madigan D,et al.Bayesian Mixed-Effects Models for Recommender Systems[C]//Proc. ACM SIGIR'99 Workshop Recommender Systems: Algorithms and Evaluation, 1999, 544-556.
[32] Ansari S. Essegaier, R. Kohli. Internet Recommendations Systems[J]. Marketing Research, 2000(8): 363-375.
[33] Canny J.Collaborative Filtering with Privacy Via Factor Analysis [C].//Proc. of the 25th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2002: 238-245.
[34] Zeng Chun,Xing Chun xiao,Zhou Li zhu.Similarity Measure and Instance Selection for Collaborative Filtering [C].//Process of the 12th International Conference on World Wide Web, 2003.
[35] Deshpande M,Karypis G.Item-based top-N Recommendation Algorithms[J]. ACM Transactionon Information Systems , 2004, 22(1): 143-177.