基于用户群组行为分析的视频推荐方法研究

李 鹏*①② 于晓洋① 孙渤禹②
①(哈尔滨理工大学测控技术与仪器黑龙江省高校重点实验室 哈尔滨 150080)
②(哈尔滨理工大学计算机科学与技术学院 哈尔滨 150080)

摘 要：该文采用权重增量及相似聚集的用户行为分析算法，为用户推荐个性化视频提供了一个有效的解决方案。方法包含 3 个主要部分，首先利用 RFM(Recentness, Frequency, Monetary amount)模型分析用户的行为，将相同行为的用户归为一组；然后结合用户的最近习惯，使用基于权重增量的 Apriori 算法挖掘用户之间的关联规则，并用向量空间模型进行相似度计算从而实现用户相似聚集；最后进行协同过滤式推荐，完成整体个性化视频推荐过程。该方法的特点是行为数据自动收集获取，避免了直接对视频大数据的处理；另外，视频推荐随着用户行为的改变而动态变化，更加符合实际情况。实验结果表明，该方法有效并且稳定，相比于单一推荐方法，在准确率、召回率等综合指标上均有明显提升。

关键词：视频推荐；行为分析；权重增量；Apriori 算法
中图分类号：TP393 **文献标识码**： A **文章编号**：1009-5896(2014)06-1485-07
DOI: 10.3724/SP.J.1146.2013.01225

Video Recommendation Method Based on Group User Behavior Analysis

Li Peng①② Yu Xiao-yang① Sun Bo-yu②
①(*Higher Educational Key Laboratory for Measuring and Control Technology, Instrumentations of Heilongjiang Province, Harbin University of Science and Technology, Harbin* 150080, *China*)
②(*School of Computer Science and Technology, Harbin University of Science and Technology, Harbin* 150080, *China*)

Abstract: This paper presents an effective solution for personalized video recommendation based on the weight increment and similar aggregation user behavior analysis algorithm. The method is implemented in three steps: first, the user behavior is analyzed using the RFM (Recentness, Frequency, Monetary amount) model, users with the same behavior are classified as a group; second, the Apriori algorithm based on weight increment is applied to mining association rules between users in line with the recent habits of users, and by using the VSM model for similarity calculation, the user similarity aggregation is realized; finally, the whole process of personalized video recommendation is completed by means of collaborative filtering. The proposed method can automatically collects user behavioral data and avoids direct video big data processing. In addition, the video recommend dynamically changes with the change of user behavior. The experiment results show that, the presented effective and stable, and the method achieves significantly increasement in precision and recall comparing with the single recommendation method.

Key words: Video recommendation; Behavior analysis; Incremental weight; Apriori algorithm

1 引言

随着互联网的迅速普及，网络传输、数据存储和视频压缩等相关技术的快速发展，来自于不同领域的各种视频数据正在以惊人的速度增长，其规模已十分庞大。例如，世界最大视频分享网站 YouTube 已经拥有超过 1.5×10^8 个视频，并且每天还有近 6.5×10^4 个新视频被上传[1]。面对如此数量级的大数据，用户想要找到自己感兴趣的视频将变成一件非常困难的事情。因此，自动的视频推荐系统成为人们迫切需求的产品，而有关推荐方法的研究也成为近年来计算机领域的一个热点研究问题，得到了国内外众多研究人员的广泛关注[2]。

协同过滤(Collaborative Filtering, CF)是目前视频推荐方法研究中被普遍接受并广泛采用的机制，也是视频推荐系统应用设计中最成功的技术[3]。

2013-08-13 收到，2013-11-08 改回
国家自然科学基金(61103149)，中国博士后科学基金(2011M500682)，黑龙江省高校青年学术骨干项目(1253G023)和哈尔滨市青年科技创新人才专项基金(2012RFQXG093)资助课题
*通信作者：李鹏 pli@hrbust.edu.cn

协同过滤的总体思想是将一位用户对于一些视频的评价与其他用户对于这些视频的评价进行比较，从而发现具有相似喜好的人，进而将这些有相似喜好人群的感兴趣视频推荐给此用户[4]。这种机制的关键在于采用何种方法来寻找并确定具有相似喜好的群体，并如何过滤出其兴趣点。目前，许多国内外研究人员已经提出了一些高水平的方法和策略来解决这一问题，并取得了巨大的进展。例如：文献[5]提出使用基于组内其它用户信息并采用联合模型来预测用户兴趣点的方法；文献[6]建立了一个用户——视频图来表示不同用户的观看信息，并通过对图的遍历从大量近邻中获取某个节点的标签从而进行推荐；文献[7]采用基于社交网络的视频推荐方法，认为社交网络中的朋友也应该具有相似的视频喜好。但是，根据统计 YouTube 用户中只有不到 40%是 FaceBook 的用户，所以这种方法只是社交网络应用发展的前期研究[8]。同时，文献[9]和文献[10]都提出了基于情感分析的视频推荐思想，前者通过判断当前用户的情绪来推荐合适的视频，后者通过识别用户的面部表情来分析用户情感并推荐适合当前情绪的视频。在他们的研究中都提出视频推荐是一种动态变化的过程，这种相似的用户群体不是一成不变的，会随着时间、情绪、感觉等诸多因素的变化而变化，是一个极其复杂的变化过程。虽然协同过滤方法取得了很大的成功，但也存在一些缺陷需要解决与完善。其中，最大的困难来自于视频推荐中两个大数据的问题，即视频数据和观看用户数据。对这两种大数据进行分类、聚类的代价都非常高，并且即使成功也很难动态变化。因此，针对视频推荐方法如何能够避免直接对两个大数据进行处理，而采用其它策略绕开大数据处理问题成为我们探索的一个方向。本文针对这些问题采用行为分析方法，将对视频和用户数据的直接操作转化为用户行为数据，用户行为数据由用户一段时间内对于视频文件的操作自动生成，并且行为数据是一种形式化的数据，便于分析处理，也适应动态的变化。

用户行为分析方法最早来源于管理学领域，通过分析客户的行为指导企业运营管理[11]。近年来，有学者将此方法的思想引入到计算机领域的研究，刘奕群等人[12]采用用户行为分析的方法对搜索引擎性能进行自动评价；陈亚睿等人[13]通过对用户行为分析模型的研究，有效遏制不可信云终端用户的侵入行为。我们认为用户对视频的点播观看行为可以反映用户对视频的兴趣态度，由此提出对一系列视频具有相似行为操作的用户应该具有相似的喜好和兴趣点的假设；本文采用的所有技术都旨在验证这个假设是否成立。

2 基于用户行为分析的视频推荐流程

本文视频推荐系统的基本流程，如图 1 所示，主要是为了用户提供个性化的视频推荐服务。用户通过界面浏览，得知视频的长短、风格、视频名称、国家地区、年代等内容标签，用户可查看视频列表并观看自己喜欢的视频，而用户事务数据库便是记录视频编号、类别风格等信息。本文通过 3 种模块阶段来呈现视频推荐的过程：

(1)用户分组模块 通过 RFM 模型对用户行为进行分析，将视频数据和观看视频客户数据转化为用户观看视频的行为操作数据，并通过日志数据对用户进行第 1 次分组；

(2)数据挖掘模块 将用户日志数据进行基于改进的权重增量的 Apriori 算法分析并取得用户频繁项的关联规则，这样可挖掘出用户在最近行为中的规则习惯；

(3)协同推荐模块 基于相似向量比对用户的相似度后，聚集相似规则用户，最后进行协同推荐，将相似比对结果做 top-N 推荐的阶段。

3 基于权重增量及相似聚集的 RFM 用户行为分析算法

3.1 基于 RFM 模型的用户行为分析

视频用户的行为分析指标是通过对用户在观看过程中的行为进行统计和分析后从中得到的一般规律所构成。通过对用户行为进行分析并且掌握用户行为的规律性，就有可能预测用户将要发生的行为来实现期望目标。分析使用视频点播服务的用户行为，是希望了解用户的特征与规律，以实现个性化推荐。用户行为分析指标主要从以下几个方面进行分析。

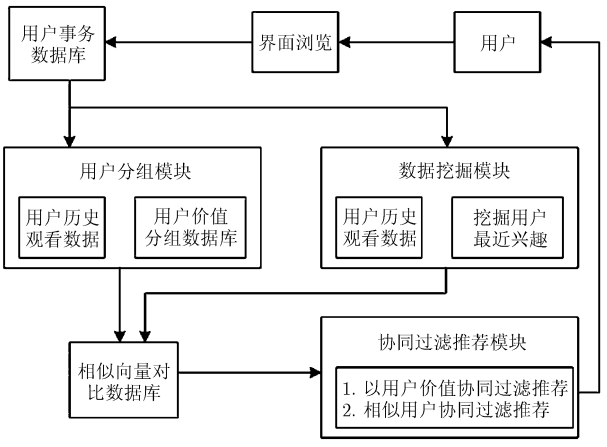

图 1 基于用户行为分析的视频推荐流程图

根据相关研究，RFM 用户数据分析的指标是由用户数据库中 3 个特殊的要素构成：最近一次消费时间(Recentness) ，消费频率(Frequency)和消费金额(Monetary Amount)，3 个要素统一到 1 个 RFM(Recentness, Frequency, Monetary amount)模型[14]。

(1)最近一次消费时间(Recentness)是指用户最后一次消费距离分析时的时间长度。当 Recentness 值较小时，用户再消费的几率比较大，因而其在最近一次消费时间特征值较高。

(2)消费频率(Frequency)是指用户在一定时间内消费该产品的次数。一般而言，当用户的消费次数越多时，该用户价值和忠诚度较高。反之，该用户价值和忠诚度较低。

(3)消费金额(Monetary Amount)是指在一段时间内，用户在此产品上花费的总金额。一般而言，当用户的消费金额越高时，其用户价值越高。

本文将对于视频的用户行为分析指标以及 RFM 的三要素做一个相对应的指标映射。如图 2 所示，我们把用户最后观看时间当作最近一次消费时间；把在一段时间内的观看频率当作消费频率；把总观看个数当作消费金额。不过，本文要将消费金额的计算方式改为计算类别文件(Itemsets)的次数，而类别文件选得越多也代表着用户会在这个类别文件上花费的时间越多，每一个类别文件就是单位金额。

本文通过行为分析可将用户分为 8 个群组，根据每一个用户的 RFM 值，我们以全部用户的 RFM 的总平均值为标准，并且以↑表示其值大于总平均值，而↓小于总平均值。利用这种表示可以分成 8 个群组(↑↑↑，↑↑↓,↑↓↑，↑↓↓，↓↑↑，↓↑↓，↓↓↑,↓↓↓)。每一位用户将其 RFM 值与平均值做一个比较，由此可以找出每一位用户的群组类型，并将每一位用户分组到符合的群组内，而系统对于每一个群组会指定不同的推荐策略。

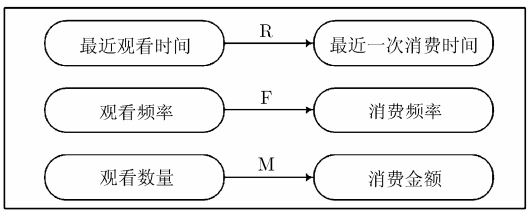

图 2 用户行为分析与 RFM 映射图

3.2 基于改进权重增量的 Apriori 算法

传统数据库由于要计算全部的观看数据，所以要获得用户的高频繁文件，势必要造成系统执行时间以及成本的增加，影响了视频推荐的即时性。并且，用户最近观看的选择也不一定会一直围绕相同的类别风格。因此，本文采用基于权重的增量式数据挖掘(Incremental Mining based on Weight, IMW)思想，从而找出用户在最近时间内的观看兴趣类别，增量式挖掘不但可以缩短数据挖掘的时间还能够动态地挖掘出用户最近习惯。

Apriori 算法作为挖掘布尔关联规则频繁项集的重要算法是迄今为止最有影响力的关联规则算法之一，其核心是基于两阶段频繁项集思想的递推算法[15]。在权重增量思想中，我们设定一个支持度阈值，只有权重支持度超过设定的支持度阈值，才能停止增量计算。随着增量计算次数的不同，所得到的结果排列也会不一样。本文通过对权重增量思想中一个参数的迭代次数阈值的设定，省去设定权重增量思想的支持度阈值以及 Apriori 算法中的最小支持度阈值，从而达到简化计算提高效率的目的。

为了计算每一类别交易项目是否是频繁集，本文定义了一个权重支持度(Weight Support, WS)，并设定 W_j 的取值 β^{j-1} 达到一个阈值时或日志交易数据取完后，即可停止计算，并且删除 WS_i^j 为零的类别文件。WS_i^j 为第 i 个类别文件在第 j 次增量观看数据中的权重支持度；C_i^j 是第 i 个类别文件出现在第 j 次增量交易的出现次数总和，j 为增量挖掘的次数，W_j 是权重值大小的计算，β^{j-1} 为 β 的 $j-1$ 次方，为一常数，其中 $\beta<1$。

$$\mathrm{WS}_i^j=\mathrm{WS}_i^{j-1}+\left(C_i^j\times W_j\right),$$
$$\mathrm{WS}_i^0=0,\ W_j=\beta^{j-1},\ j=1,2,\cdots,n \tag{1}$$

本文方法是将权重增量的思想加入到 Apriori 算法里并进行改进，从而求得研究中理想的规则，以下是描述挖掘规则的步骤：

步骤 1　假设先取观看数据库内的最后 n 笔交易，并计算每一项集类别 i 的次数值。

步骤 2　以式(1)计算每一个类别的 WS 值，判断 W_j 的取值 β^{j-1} 的值是否达到一个阈值或日志交易数据取完。

步骤 3　如果不符合就再取下一个 n 笔观看数据，并且再重新计算 WS 值，直到 W_j 的取值 β^{j-1} 达到一个阈值或日志数据取完后，停止计算。

步骤 4　删除 WS_i^j 为零或不足最小支持度 MS 值的项集项目，并通过一项集类别的高频繁项目，组合成二项候选集合，重复上述的方式计算每一个

二项集类别 i。

步骤 5 二项候选集合依据之前的一项类别集合所做的增量次数($j=1, 2,\cdots, n$)。接着删除 WS_i^j 为零或不足最小支持度 MS 值的集合。

步骤 6 最后剩下的二项集类别将视为用户的最近习惯规则(Recent behavior rules, Rbr)。

$$\mathrm{Rbr}=\left\{[i\to j]\middle\|\mathrm{MS}\ \le\ \mathrm{WS}([i\ \to j]),\ \ \beta^{j-1}<0.1\right\} \quad (2)$$

式(2)中，将 WS 大于 MS 的二项类别文件归为最近习惯和兴趣规则。其中 $[i\to j]$ 为可能类别项目的表示。

3.3 基于 VSM 模型的用户相似聚集

本文通过向量空间模型(Vector Space Model, VSM)对用户进行相似度计算，并且依据用户最近习惯和兴趣得出的规则来做用户聚类，对相似用户进行再一次聚集。其目的是为了聚集相似类别项目的用户，找出用户间更加相似的群组，达到真正协同过滤方法下分享信息的作用。定义如下：

$\mathbf{SM}(X)$ 为用户 U_x 最近习惯规则 Rbr 展开结合成的相似矩阵(Similar Matrix, SM), Rbr 为最近习惯规则 $[i\to j]$ 可能类别项目集合，式(3)所示。

$$\mathbf{SM}(X)=\left[sm_{ij}\right]m\times n,\ \ m,n=1,2,\cdots,$$
$$p,sm_{ij}=\begin{cases}1, & [i\to j]\in \mathrm{Rbr}\\ 0, & \text{其它}\end{cases} \quad (3)$$

接下来进行相似向量的计算，相似向量的定义如下：

$$\mathbf{SV}(X)=\left[sm_{12}\ \ sm_{13}\cdots sm_{1n}\ \ sm_{23}\ \ sm_{24}\cdots sm_{2n}\cdots sm_{m(m+1)}\cdots sm_{mn}\right] \quad (4)$$

求得相似向量后，就可进行每一用户之间的相似度对比。

本文采用空间向量模型进行用户之间的相似度对比。在向量空间模型中，两位用户 D_1 和 D_2 之间的行为相似度 $\mathrm{Sim}(D_1,\ D_2)$ 常用向量之间夹角的余弦值表示，如式(5)：

$$\mathrm{Sim}(D_1,D_2)=\cos\theta=\frac{\sum_{k=1}^{n}W_{1k}\times W_{2k}}{\sqrt{\left(\sum_{k=1}^{n}W_{1k}^2\right)\left(\sum_{k=1}^{n}W_{2k}^2\right)}} \quad (5)$$

其中 W_{1k}, W_{2k} 分别表示文本 D_1 和 D_2 第 K 个特征项的权值，$1\le k\ \le N$。若 n 维向量是 $\boldsymbol{V}\{v_1,v_2,\cdots,v_n\}$，则模 $|v|=\mathrm{sqrt}\ (\ v_1\times v_1+v_2\times v_2+\cdots+v_n\times v_n\ )$，两个向量的乘积 $\boldsymbol{s}\times\boldsymbol{z}=z_1\times s_1+z_2\times s_2+\cdots+z_n\times s_n$，这时两个向量的相似度为 $(\boldsymbol{s}\times\boldsymbol{z})/(|\boldsymbol{s}|\times|\boldsymbol{z}|)$。

通过 VSM 模型，就可以对群组中用户的最近习惯规则向量表示做相似度计算，向量中的每一个元素都作为向量的特征项，对同一分组中的两两用户做相似度计算 ，并以相似度作为系数，对所要推荐给其他用户的视频做推荐度分析。若两个用户的类别相似向量的相似度高，那么就将一个用户的视频以高比例的数量推荐给其他用户，若两个用户的类别相似向量的相似度低，就将一个用户的视频以低比例的数量推荐给其他用户。

3.4 基于协同过滤的视频推荐

本文采用协同过滤的方式向用户进行视频推荐。在分组模块中，利用 RFM 模型进行用户行为分析，并分成 8 个用户群组。网站可以通过这种方式针对价值度不同的用户进行策略性推荐。此推荐方法将所有用户选择的视频作为推荐的视频，如果用户属于高价值群组，推荐的视频就以挖掘结果做协同过滤式推荐；反之，如果用户属于低价值群组，推荐的视频就依挖掘结果随机推荐不相关视频。例如：用户 U_1 属于高价值群组，其挖掘结果为 $\{A\to C\}$ 和 $\{A\to D\}$ 两种规则，我们将所有用户选择有关 A, C, D 的视频，推荐给用户 U_1，达到精确个性化视频推荐。如果用户 U_1 是属于低价值群组内，我们不只推荐相关的视频，还要增加其它类别视频，道理类似于多做行销广告的策略，好让这个群组的用户能够有机会多登陆使用并且增加历史记录数据，这样才能更好地为用户服务。

针对同一群组内的用户，我们进行了组内相似用户聚类，聚集了同一群组内与其他用户最相近风格的用户。本方法是利用 RFM 模型分类后，在相同群的其他用户所点选的视频来进行相互推荐。这种方式的目的是经由第 2 次的分类聚集，可以得出更接近用户习惯和兴趣分类，其做法就是将其他用户所选择的视频，依据之前用户的喜好类别不重复地推荐给用户，达到协同过滤式信息共享的结果。

假设 U_1 的同组其他相似同喜好用户 U_2 和 U_3，可以知道3位用户所选择的视频编号(其中 vt 代表第 t 个视频)。先把 U_1 与 U_2 在相同类别中的，不重复地推荐给 U_1，例如：U_2 将{A:v1,v2, B:v6, C:v9}推荐给 U_1。而 U_3 相同的类别也推荐给 U_1，例如：U_3 将{A:v2,v4, B:v6,v8}推荐给 U_1。这样，就会综合 U_2 与 U_3 两者的结果不重复地推荐给 U_1，即{A:v1,v2, v4, B:v6,v8, C:v9}。

4 实验验证与分析

针对视频推荐方法的评价是一个比较困难的问题。由于视频推荐的对象是人，因而对视频喜好的选择因人而异，甚至在不同时间、不同环境下同一个人的选择也存在差异。因此，人们无法构建统一

的公共数据集来衡量各种方法之间的优劣，绝大多数的研究只能通过组织一定数量的用户对自己的方法与基本方法进行评价，以验证自己提出的策略或所采用的技术是否有效且稳定。我们也采用这种实验策略，通过组织本中心的部分人员作为实验者，对本文所采用的技术进行评价。

4.1 实验平台构建

为验证本文方法的有效性和稳定性，本文搭建了一个实验平台。实验数据分为 5 种类型共有 400 个视频，其中 250 个为训练语料，剩余 150 个为测试语料，每种分类的前 50 个作为训练语料，后 30 个作为测试语料。本实验共有 15 位实验者在 30 天内生成了 1733 条日志数据。为了证明本文方法的有效性与稳定性、本文构造了 4 种方法，分别用以验证 RFM 模型，权重增量以及相似聚集 3 种技术在推荐过程中分别所起的作用，4 种方法分别表示如下：

方法 1：　使用权重增量与 RFM 模型。

方法 2：　使用权重增量与 RFM 模型及相似聚集。

方法 3：　不使用权重增量但使用 RFM 模型及相似聚集。

方法 4：　使用权重增量但是不加入任何分组方法。

本实验采用准确率、召回率和 F 值这 3 个指标来衡量实验方法的有效性，为了计算实验数据的准确率和召回率，评价指标的计算公式如式(6)，式(7)，式(8)所示。本文将实验和分类问题的混淆矩阵相结合，从而更好地描述系统的性能，如表 1 所示。其中 TP 表示的是方法推荐的并且用户真实喜欢的视频数，FP 表示方法推荐的但不是用户喜欢的视频数，FN 表示方法没有推荐但是用户实际喜欢的视频数，而 TN 则是方法既没有推荐而且用户也不喜欢的视频数。

表 1 分类混淆矩阵

	用户实际喜欢的视频	用户实际不喜欢的视频
方法推荐的视频	TP	FP
方法没有推荐的视频	FN	TN

$$\text{准确率}(\mathrm{Pr}) = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FP})} \tag{6}$$

$$\text{召回率}(\mathrm{Re}) = \frac{\mathrm{TP}}{(\mathrm{TP}+\mathrm{FN})} \tag{7}$$

$$F\text{值}(F) = \frac{2\times \mathrm{Pr}\times \mathrm{Re}}{\mathrm{Pr}+\mathrm{Re}} \tag{8}$$

4.2 实验结果与分析

通过对 15 名实验者观看视频的行为数据进行采集并分析，采用构建的 4 种方法分别进行视频推荐，得到了如表 2 所示的实验结果。

表 2 用户的权重增量+RFM+相似聚集与其它方法的评价指标对比

	方法1	方法2	方法3	方法4
平均准确率	0.64	0.76	0.52	0.56
平均召回率	0.67	0.79	0.65	0.64
平均 F 值	0.65	0.78	0.58	0.60
时间复杂度	$O(n)$	$O(n^2)$	$O(n^2)$	$O(\lg n)$

方法 2 与其它方法的比较可以说明，单纯只考虑 RFM 模型分组在推荐的过程中可能会出现较多其他用户的推荐，因此可能推荐一些非用户喜好或兴趣的视频，所以，本实验设计证明两次分组效果优于单独的一次分组。而方法 3 不使用增量挖掘技术，推荐时会不管用户的喜好，任意推荐其他用户所点选的视频，造成推荐比较杂乱，所以用户对这种混乱的推荐可能不喜欢，因此推荐后的准确率明显低于使用增量挖掘的方法，证明权重增量挖掘的重要性。另外，方法 4 不使用分组技术，其平均准确率为 56%，明显低于方法 2 使用分组技术的准确率，因此分组技术更能让用户可以得到想要的，而不是一堆视频，让用户不知道怎么选。以上实验数据证明了方法 2 的有效性，但还需要验证方法的稳定性。

图 3 显示了 15 位用户的准确率的分布。可以看出，用户 7 在推荐方法 2 的准确率最高为 82%，而最低是用户 2 使用推荐方法 2 的准确率也有 68%。其中用户在使用方法 2 时所得到的大多数的准确率数值明显高于其它 3 种方法所得到的准确率，其中方法 1 和方法 4 在准确率的稳定性相对较差。

图 4 显示了 15 位用户的召回率的分布。用户 4 使用方法 2 所得到的召回率最高为 84%，而最低的是用户 12 的召回率也有 72%。其中用户在使用方法 2 时所得到的大部分的召回率数值明显高于其它 3 种方法所得到的召回率，其中方法 1 和方法 3 在召回率的稳定性相对较差。

本文对文中视频协同推荐框架下所涉及的 3 种算法分别进行了时间复杂度分析：基于 RFM 模型的用户行为分析算法(RFM)本质是一种匹配算法，算法对用户 3 种行为元素进行采样，并与可能形成的 8 种情况进行匹配，将用户进行群组集聚。这种匹配算法没有循环存在，因此其时间复杂度为常数；

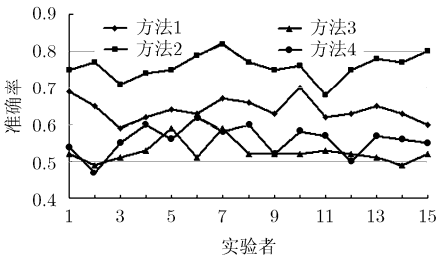

图 3 4 种方法的准确率分布图

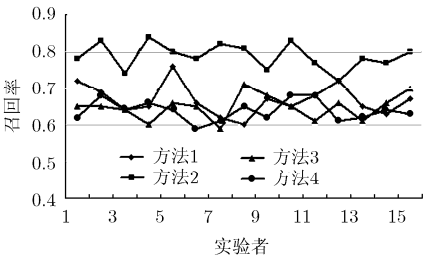

图 4 4 种方法的召回率分布图

基于改进权重增量的 Apriori 算法(IMW)本质是一种递归算法，算法主要对于用户的近期观看行为规则进行增量式的更新，匹配算法的时间复杂度最小为 $O(n)$，采用折半查找时间复杂度最大为 $O(\lg n)$；向量空间模型(VSM)是一种普遍使用的高效相似度计算模型，VSM 内积计算的时间复杂度是 $O(n)$，待推荐的用户要与已知用户集分别进行相似度计算，其时间复杂度也为 $O(n)$。因此，基于 VSM 模型的用户相似聚集算法(similarity)的时间复杂度为 $O(n^2)$。通过以上分别对 3 种算法的时间复杂度分析，可以对实验中所验证 4 种方法的时间复杂度进行对比，具体如表 2 所示。可以看到，方法 1 的时间复杂度最小为 $O(n)$，方法 2 和方法 3 的时间复杂度最大为 $O(n^2)$，影响时间复杂度的主要因素是采用 VSM 模型进行用户相似聚集。但是从其它指标上综合考虑，此算法对于提升视频推荐效果确实起到了重要的作用。从代价上考虑，在实际应用系统中算法的时间复杂度为 $O(n^2)$是可以被接受的，如支持向量机(SVM)算法被广泛地应用于各种实际系统开发之中，其时间复杂度即为 $O(n^2)$。

通过对以上数据的分析可以看到，方法 2 利用权重增量及相似聚集的 RFM 模型推荐方法，能够更好地发现用户的喜好，从而相比其它基本方法具有更好的推荐能力，具有较高的准确率，并在一定程度上也表现出了方法的稳定性。因此，可以证明本文研究方法中所涉及的 3 种技术，即权重增量挖掘、组内用户相似聚集以及基于 RFM 模型的用户行为分析均对视频推荐具有正向推动，是一种有效的手段。另外，也进一步证明了本文先前的假设是成立的，即对一系列视频具有相似行为操作的用户应该具有相似的喜好和兴趣点。

5 结束语

本文首先通过 RFM 模型将价值或者行为相同用户归为同一群组，结合用户最近习惯和行为，采用 Apriori 算法来挖掘关联式规则；然后用相似向量矩阵计算所有用户之间的相似度关系，进行相似聚集；最后利用协同过滤式推荐方法给用户进行视频推荐，从而完成个性化推荐的整个过程。本文通过实验结果验证了此推荐方法的有效性和稳定性。结合RFM模型及相似聚集推荐比单纯只使用RFM模型分组方式效果好，利用权重增量挖掘与分组方式实验结果表明，能够推荐给用户更准确的喜好视频。而整体上，本实验的准确率高达 76%，比其它推荐方法高出 16.2%～32.5%，召回率高达 79%，比其它推荐方法高出 15.1%～18.9%。综合上述实验结果，可以证明本文所采用的 3 种技术相结合的方法是一种行之有效的视频推荐策略，基本达到了预期的效果。

本文的主要贡献在于提出了采用用户行为分析的方法对视频进行推荐，目前还没有查阅到同样采用行为分析进行视频推荐的相关文献。通过自动采集用户观看视频的行为数据，并通过技术手段分析这些数据找到具有相同喜好的用户，进而进行协同推荐。行为数据可以实现动态实时采集，行为数据属于形式化数据，其处理难度小、速度快，从而可以实现及时更新，同时也避免了以巨大代价对视频大数据进行的直接处理。在视频推荐的实际应用中，推荐的及时性往往比推荐方法的准确性更重要，因此对其应用研究不能仅着眼于算法的复杂化，而相反应该寻找简单、稳定的策略。在今后的研究中，我们将继续深入探索基于行为分析的视频推荐方法，积极研究用户深层次行为属性特点，丰富行为模式内涵。

参 考 文 献

[1] SKrishnapp, D K, Zink M, and Griwodz C. Cache-centric video recommendation: an approach to improve the efficiency if YouTube caches[C]. Preceedings of the 4th ACM Multimedia System Conference, Oslo, 2013: 261–270.

[2] Zhao Xiao-jian, Yuan Jin, and Wang Meng. Video recommendation over multiple information sources[J]. *Multimedia Systems*, 2011, 19(1): 3–15.

[3] De V J, Degrande N, and Verhoeyen M. Video content recommendation: an overview and discussion on technologies

and business models[J]. *Bell Labs Technical Hournal*, 2011, 16(2): 235–250.

[4] Park J, Lee S, and Kim K. Online video recommendation through tag-cloud aggregation[J]. *IEEE MultiMedia,* 2011, 18(1): 78–87.

[5] Su Chun-rong, Li Yu-wei and Zhang Rui-zhe. An adaptive video program recommender based on group user profiles[J]. *Smart Innovation, Systems and Technologies*, 2013, 21(2): 499–509.

[6] Ozturk G and Kesim C N. A hybrid video recommendation system using a graph-based algorithm[J]. *Lecture Notes in Computer Science*, 2011, 6704: 406–415.

[7] Silveira D, Alessandro, and Wives L K. POI enhanced video recommender system using collaboration and social networks[C]. Preceedings of the 8th International Conference on Web Information Systems and Technologies, Valencia, 2012: 717–722.

[8] Ma Xiao-qiang, Wang Hai-yang, and Li Hai-tao. Exploring sharing patterns for video recommendation on YouTube-like social media[J]. *Multimedia Systems*, 2013, DOI: 1007/s00530-013-0309-1.

[9] Niu Jian-wei, Zhao Xiao-ke, Zhu Li-ke, *et al.*. Affivir: an affect-based internet video recommendation system[J]. *Neurocomputing*, 2013, 120: 422–433.

[10] Zhao Si-cheng, Yao Hong-xun, and Sun Xiao-shuai. Video classification and recommendation based on affective analysis of viewers[J]. *Neurocomputing*, 2013, 119: 101–110.

[11] Rapach D E and Wohar M E. Forecasting the recent behavior of US business fixed investment spending: an analysis of competing models[J]. *Journal of Forecasting*, 2007, 26(1): 33–51.

[12] 刘奕群, 岑荣伟, 张敏. 基于用户行为分析的搜索引擎自动性能评价[J]. 软件学报, 2008, 19(11): 3023–3032.
Liu Yi-qun, Cen Rong-wei, and Zhang Min. Automatic search engine performance evaluation based on user behavior analysis[J]. *Journal of Software*, 2008, 19(11): 3023–3032.

[13] 陈亚睿, 田立勤, 杨扬. 云计算环境下基于动态博弈论的用户行为模型与分析[J]. 电子学报, 2011, 39(8): 1818–1823.
Chen Ya-rui, Tian Li-qin, and Yang Yang. Model and analysis of user behavior based on dynamic game theory in cloud computing[J]. *Acta Electronica Sinica*, 2011, 39(8): 1818–1823.

[14] Chen Toly. The RFM-FCM approach for customer clustering[J]. *International Journal of Technology Intelligence and Planning*, 2012, 8(4): 358–373.

[15] Awadalla M H and Elfar S G. Aggregate function based enhanced apriori algorithm for mining association rules[J]. *International Journal of Computer Science Issues*, 2012, 9(3): 277–287.

李　鹏：　男，1978 年生，教授，硕士生导师，研究方向为网络信息处理、机器学习、人工智能.

于晓洋：　男，1962 年生，教授，博士生导师，研究方向为图像加密与隐藏、视觉三维检测.