



(12) 发明专利

(10) 授权公告号 CN 103377242 B

(45) 授权公告日 2016.06.22

(21) 申请号 201210127442.3

(22) 申请日 2012.04.25

(73) 专利权人 TCL 集团股份有限公司

地址 516001 广东省惠州市鹅岭南路6号
TCL 工业大厦8楼技术中心

(72) 发明人 董延平 汪灏泓

(74) 专利代理机构 深圳市君胜知识产权代理事
务所 44268

代理人 刘文求

(51) Int. Cl.

G06F 17/30(2006.01)

H04N 21/258(2011.01)

H04N 21/462(2011.01)

(56) 对比文件

CN 101702653 A, 2010.05.05,

CN 101945400 A, 2011.01.12,

CN 101094335 A, 2007.12.26,

US 2003/0101451 A1, 2003.05.29,

Chin-Feng Lai 等.CPRS: A cloud-based
program recommendation system for digital
TV platforms.《Future Generation Computer
Systems》.2011,第27卷(第6期),823-835.

Yi-Hung Wu 等.Enabling Personalized
Recommendation on the Web Based on User
Interests and Behaviors.《Proceedings
Eleventh International Workshop on Research
Issues in Data Engineering》.2010,17-24.

审查员 邱川

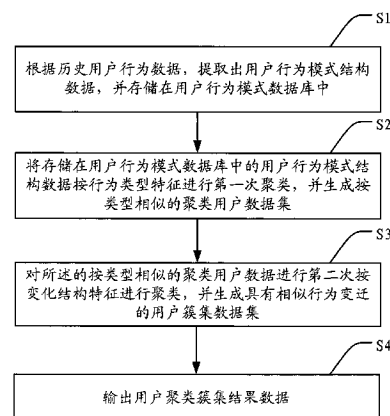
权利要求书2页 说明书8页 附图4页

(54) 发明名称

用户行为分析方法、分析预测方法及电视节目
推送系统

(57) 摘要

本发明公开了一种用户行为分析方法、分析预测方法及电视节目推送系统,首先对历史用户行为数据进行分析,提取出用户行为模式结构数据,并存储在用户行为模式数据库中;然后将用户行为模式结构数据按行为类型特征进行第一次聚类,生成按类型相似的聚类用户数据集;再对按类型相似的聚类用户数据进行第二次按变化结构特征进行聚类,生成具有相似行为变迁的用户簇集数据集;最后输出用户聚类结果数据。因为第二次聚类考虑了用户行为的变迁时序特征,其中蕴含着单纯统计数据中没有的用户行为变迁信息,使得对用户的描述会更加的完备,且最后得到的聚类结果集可以很方便的进行用户之间的协同分析,并在电视节目推广领域,为用户推送潜在兴趣节目。



1. 一种用户行为分析方法,其特征在于,包括以下步骤:

ST1、根据历史用户行为数据,提取出用户行为模式结构数据,并存储在用户行为模式数据库中;

ST2、将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户数据集;

ST3、对所述的按类型相似的聚类用户数据进行第二次按变化结构特征进行聚类,并生成具有相似行为变迁的用户簇集数据集;

ST4、输出用户聚类簇集结果数据;

所述步骤ST1中,用户行为模式结构数据为 $D1=(U,P)$ 其中, $U=\{\text{用户1、用户2}\cdots\cdots\text{用户n}\}$ 代表用户集合, $P=P(c,s,n,m,k)$,其中, c 为行为类型, s 为行为起始时间, n 为行为间隔时间, m 为行为出现周期次数, k 为行为时间长度;

所述步骤ST2中,所述对所述的按类型用户行为数据进行第一次聚类具体在于:

定义聚类的用户行为类型相似度表达式为 $\text{sim}(i,j)=\frac{C_i \cap C_j}{C_i \cup C_j}$;其中, C_i 、 C_j 表示 U 中第 i 、 j

个用户的行为模式类集合;

对结构数据 $D1=(U,P)$, $P(c,s,n,m,k)$ 的行为类型 c 按所述的用户行为类型相似度表达式进行聚类分析;

所述步骤ST3具体在于:

定义所述按类型用户行为数据集为 $D2=(U,S)$,其中 $U=\{\text{用户1、用户2}\cdots\cdots\text{用户n}\}$; $S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$, s_{ij} 为第 i 个用户的第 j 个行为模式起点, n_{ij} 第 i 个用户的第 j 个行为为模式周期, m_{ij} 第 i 个用户的第 j 个行为为周期数;

定义则聚类的变化特征相似度表达式为 $\text{sim}(i,j)=\frac{\sum w_{ij}}{|C_i \cup C_j|}$

其中, w_{ij} 的值为1时代表用户 i 与 j 用户行为模式类型相同且次序相同,否则为0; $|C_i \cup C_j|$ 为用户 i 与用户 j 存在的类型元素总个数;

对结构数据 $D2=(U,S)$, $S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$,按所述的变化特征相似度表达式进行聚类分析。

2. 根据权利要求1所述的用户行为分析方法,其特征在于,所述的聚类分析基于MST的聚类分析,具体包括以下步骤:

STA、用相似度表达式,算出所有点之间的权值;

STB、使用MST聚类构造算法,生成最小生成树;

STC、设定可容忍差异权值,大与或等于其值的边定义为边界边;

STD、打断所有边界边;根据打断数量来确定最终生成的聚类数。

3. 根据权利要求2所述的用户行为分析方法,其特征在于,所述步骤STB中基于MST聚类的最小生成树构造算法为普里姆算法或克鲁斯卡尔算法。

4. 一种用户行为分析预测方法,其特征在于,所述方法包括以下步骤:

STA1、建立用户行为分析预测系统,所述用户行为分析预测系统,存储有按变化结构特征用户行为数据集,所述的按变化结构特征用户行为数据集,采用如权利要求1至3任一项

所述的用户行为分析方法生成；

STB1、收集当前用户行为数据；

STC1、将所述用户行为数据输入所述的用户行为分析预测系统，所述的用户行为分析预测系统生成按变化特征用户行为进行聚类的数据聚类簇集，并通过聚类结果中其它用户预测用户行为变迁的可能性。

5. 一种电视节目推送系统，其特征在于，所述的系统包括：

智能电视终端，用于播放节目与采集即时用户行为数据；

用户行为聚类分析引擎，用于根据所述的即时用户行为数据进行包括按类型、按变化结构特征的两次聚类分析，将具有相同历史行为变迁的用户分为一类形成簇集，作为后期推荐数据集；

节目推送云端服务器，用于根据聚类结果集推送用户潜在兴趣节目；

所述的智能电视终端包括：用户即时行为数据采集与传输装置，用于与所述的用户行为聚类分析引擎建立网络连接，采集与发送用户即时行为数据；

所述的用户行为聚类分析引擎包括：数据接收装置，用于与所述的智能电视终端建立网络连接，接收与存储所述的发送用户即时行为数据；数据聚类分析装置，用于根据如权利要求1至3任一项所述的用户行为分析方法对所述的用户即时行为数据进行聚类分析，将具有相同历史行为变迁的用户分为一类形成簇集，作为后期推荐数据集；

所述的节目推送云端服务器包括：用户潜在兴趣节目挖掘装置，所述的挖掘装置与所述的用户行为聚类分析引擎建立网络连接，并采用如权利要求4所述的用户行为分析预测方法，根据挖掘分析出用户聚类集，并输出挖掘结果数据集；节目推送装置，用于与所述的智能电视终端建立网络连接，直接推送用户潜在兴趣节目或再经过协同过滤对潜在兴趣节目进行排序然后再进行推送。

6. 如权利要求5所述的电视节目推送系统，其特征在于，

所述节目推送装置经过协同过滤对潜在兴趣节目进行排序然后再进行推送的具体步骤如下：

STA2、列举出潜在兴趣节目；

STB2、根据节目推荐度对潜在兴趣节目进行排序；

STC2、将排序后的潜在兴趣节目推荐给用户。

7. 如权利要求6所述的电视节目推送系统，其特征在于，

所述步骤STB2中节目推荐度的计算方法如下：

$$P = \frac{\sum_{j=1}^n \text{Sim}_i * I_j}{\sum_{j=1}^n \text{Sim}_i} ;$$

其中，节目推荐度P的取值范围在0到1之间；

j表示用户数量，其取值范围在1到n之间；

I_j 表示用户j是否观看过，当观看时 I_j 为1，否则为0；

Sim_{ij} 表示目标用户i与比对对象j之间的相似度。

用户行为分析方法、分析预测方法及电视节目推送系统

技术领域

[0001] 本发明涉及数据挖掘技术领域,特别涉及一种用户行为分析方法、分析预测方法及电视节目推送系统。

背景技术

[0002] 目前,大部分算法都用的是统计数据进行初步数据处理,这样的数据处理用户行为数据会丢失行为时序与周期性特点,正是因为这些数据特性的丢失,会造成用户描述的准确性不理想。而且同一个用户的行为有太多的不可预知性,如果仅仅针对用户的统计数据进行分析,则很难完整的预测用户下次的行为,例如:同为周一用户统计规律会去看A类节目,由于未知的原因去看了B类节目,但你根据统计的描述无法发现隐藏在用户观看历史中的行为变迁信息,故是不会去推荐B类节目,从而无法为用户提供人性化的消费品和服务。

[0003] 因此,现有技术还有待改进和提高。

发明内容

[0004] 本发明的目的在于提供一种用户行为分析方法、分析预测方法及电视节目推送系统,以解决现有技术中对用户行为分析不准确、会丢失行为时序与周期性特点,造成用户行为描述的准确性较差的问题。所述用户行为分析方法基于变化结构的数据分析方法,所述分析预测方法基于所述用户行为分析方法,所述电视节目推送系统基于所述分析预测方法。

[0005] 为了达到上述目的,本发明采取了以下技术方案:

[0006] 一种用户行为分析方法,其中,包括以下步骤:

[0007] ST1、根据历史用户行为数据,提取出用户行为模式结构数据,并存储在用户行为模式数据库中;

[0008] ST2、将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户数据集;

[0009] ST3、对所述的按类型相似的聚类用户数据进行第二次按变化结构特征进行聚类,并生成具有相似行为变迁的用户簇集数据集;

[0010] ST4、输出用户聚类簇集结果数据。

[0011] 所述的用户行为分析方法,其中,所述步骤ST1,对历史用户行为数据进行分析,提取出用户行为模式结构数据,并存储在用户行为模式数据库中,用户行为模式结构数据为 $D1=(U,P)$ 其中, $U=\{\text{用户1、用户2} \dots \text{用户n}\}$ 代表用户集合, $P=P(c,s,n,m,k)$,其中, c 为行为类型, s 为行为起始时间, n 为行为间隔时间, m 为行为出现周期次数, k 为行为时间长度。

[0012] 所述的用户行为分析方法,其中,所述步骤ST2,将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户

数据集中,所述对所述的按类型用户行为数据进行第一次聚类具体在于:

[0013] 定义聚类的用户行为类型相似度表达式为 $\text{sim}(i, j) = \frac{C_i \cap C_j}{C_i \cup C_j}$;其中, C_i 、 C_j 表示U中

第i、j个用户的行为模式类集合;

[0014] 对结构数据 $D1=(U, P)$, $P(c, s, n, m, k)$ 的行为类型c按所述的用户行为类型相似度表达式进行聚类分析。

[0015] 所述的用户行为分析方法,其中,在所述步骤ST3,对所述的按类型相似的聚类用户数据进行第二次按变化结构特征进行聚类,并生成具有相似行为变迁的用户簇集数据集中,具体在于:

[0016] 定义所述按类型用户行为数据集为 $D2=(U, S)$,其中 $U=\{\text{用户1、用户2} \dots \dots \text{用户n}\}$; $S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$, s_{ij} 为模式起点, n_{ij} 为模式周期, m_{ij} 为周期数;

[0017] 定义则聚类的变化特征相似度表达式为 $\text{sim}(i, j) = \frac{\sum w_{ij}}{|C_i \cup C_j|}$

[0018] 其中, w_{ij} 的值为1时代表用户i与j用户行为模式类型相同且次序相同,否则为0; $|C_i \cup C_j|$ 为用户i与用户j存在的类型元素总个数;

[0019] 对结构数据 $D2=(U, S)$, $S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$,按所述的变化特征相似度表达式进行聚类分析。

[0020] 所述的用户行为分析方法,其中,所述的聚类分析为基于MST的聚类分析。所述基于MST的聚类分析具体包括以下步骤:

[0021] STA、用相似度表达式,算出所有点之间的权值;

[0022] STB、使用MST聚类构造算法,生成最小生成树;

[0023] STC、设定可容忍差异权值,大与或等于其值的边定义为边界边;

[0024] STD、打断所有边界边;根据打断数量来确定最终生成的聚类数。

[0025] 所述的用户行为分析方法,其中,所述步骤STB中基于MST聚类的最小生成树构造算法为普里姆算法或克鲁斯卡尔算法。

[0026] 一种用户行为分析预测方法,其中,所述方法包括以下步骤:

[0027] STA1、并建立用户行为分析预测系统,所述用户行为分析预测系统,存储有按变化结构特征用户行为数据集;

[0028] STB1、收集当前用户行为数据;

[0029] STC1、将所述用户行为数据输入所述的用户行为分析预测系统,所述的用户行为分析预测系统生成按变化特征用户行为进行聚类的数据聚类簇集,并通过聚类结果中其它用户预测用户行为变迁的可能性。

[0030] 一种电视节目推送系统,其中,所述的系统包括:

[0031] 智能电视终端,用于播放节目与采集即时用户行为数据;

[0032] 用户行为聚类分析引擎,用于根据所述的即时用户行为数据进行包括按类型、按变化结构特征的两次聚类分析,将具有相同历史行为变迁的用户分为一类形成簇集,作为后期推荐数据集;

[0033] 节目推送云端服务器,用于根据聚类结果集推送用户潜在兴趣节目。

[0034] 所述的电视节目推送系统,其中,

[0035] 所述的智能电视终端包括：用户即时行为数据采集与传输装置，用于与所述的用戶行为聚类分析引擎建立网络连接，采集与发送用户即时行为数据；

[0036] 所述的用户行为聚类分析引擎包括：数据接收装置，用于与所述的智能电视终端建立网络连接，接收与存储所述的发送用户即时行为数据；数据聚类分析装置，用于根据所述的用户行为分析方法对所述的用户即时行为数据进行聚类分析，生成按变化结构特征用户行为聚类的簇集；

[0037] 所述的节目推送云端服务器包括：用户潜在兴趣节目挖掘装置，所述的挖掘装置与所述的用户行为聚类分析引擎建立网络连接，并采用所述的用户行为分析预测方法，根据所述的按变化结构特征用户行为数据集对用户潜在收看节目进行挖掘分析，并输出挖掘结果数据集；节目推送装置，用于与所述的智能电视终端建立网络连接，并推送用户潜在兴趣节目。

[0038] 所述的电视节目推送系统，其中，

[0039] 所述节目推送装置经过协同过滤对潜在兴趣节目进行排序然后再进行推送的具体步骤如下：

[0040] STA2、列举出潜在兴趣节目；

[0041] STB2、根据节目推荐度对潜在兴趣节目进行排序；

[0042] STC2、将排序后的潜在兴趣节目推荐给用户。

[0043] 所述的电视节目推送系统，其中，

[0044] 所述步骤STB2中节目推荐度P的计算方法如下：

$$P = \frac{\sum_{j=1}^n Sim_{ij} * I_j}{\sum_{j=1}^n Sim_{ij}} ;$$

[0046] 其中，P的取值范围在0到1之间；

[0047] j表示用户数量，其取值范围在1到n之间；

[0048] I_j 表示用户j的观看可能性，当观看时 I_j 为1，否则为0；

[0049] Sim_{ij} 表示目标用户i与比对对象j之间的相似度

[0050] 有益效果：

[0051] 本发明提供的用户行为分析方法、分析预测方法及电视节目推送系统，通过在变化结构中蕴含着单纯统计数据中没有的用户行为变迁信息，使得对用户的描述会更加的完备，而且可以很方便的进行用户之间的协同分析(即协同过滤)，通过近似用户的变化来预测用户的可能变化。

附图说明

[0052] 图1为本发明的用户行为分析方法实施例的流程图。

[0053] 图2为图1中的用户行为模式的数据结构的结构框图。

[0054] 图3为图2中一具体实施例形成的向量差异矩阵图。

[0055] 图4(a)为图3形成的无向图。

- [0056] 图4(b)为图4(a)的最小生成树。
- [0057] 图4(c)是4(b)“可容忍差异”权值大于0.3的聚类结果图。
- [0058] 而图4(d)是4(b)“可容忍差异”权值大于0.27的聚类结果图。
- [0059] 图5为本发明的用户行为分析预测方法的流程图。
- [0060] 图6为本发明的智能电视节目推送系统的结构框图。

具体实施方式

[0061] 为使本发明的目的、技术方案及效果更加清楚、明确,以下参照附图并举实例对本发明进一步详细说明。

[0062] 请参阅图1,其为本发明的用户行为分析方法实施例的流程图,如图所示,所述方法包括以下步骤:

[0063] S1、根据历史用户行为数据,提取出用户行为模式结构数据,并存储在用户行为模式数据库中;

[0064] S2、将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户数据集;

[0065] S3、对所述的按类型相似的聚类用户数据进行第二次按变化结构特征聚类,并生成具有相似行为变迁的用户簇集;

[0066] S4、输出用户聚类簇集结果数据。

[0067] 下面分别针对上述步骤进行详细说明:

[0068] 我们知道,可以从历史用户行为中推测出用户行为模式,因此,在第一步骤S1中,通过对历史用户行为数据进行分析,提取出用户行为数据对应的用户行为模式的行为模式数据结构,在行为模式数据结构数据表中记录所述用户行为数据,并存储在数据库中。例如:记录用户过往电视节目观看历史(即分析历史用户行为数据),集中发掘出主要的用户观看模式,建立观看行为模式结构数据表,在所述观看行为模式结构数据表中记录所述用户观看行为数据,从而为后期用户行为模式识别选出关键点数据。在本发明第一实施例中,用户行为模式结构数据为 $D1=(U,P)$ 其中, $U=\{\text{用户1、用户2} \dots \text{用户n}\}$ 代表用户集合, $P=P(c,s,n,m,k)$,其中, c 为行为类型, s 为行为起始时间, n 为行为间隔时间, m 为行为出现周期次数, k 为行为时间长度,其结构如图2所示。这种数据结构记录了时间起点,这样可以获取到每一个同期性行为的起点,从而可以分析得到每个行为的时序特征。通过 n 和 m 来获取用户的同期特性与行为结束点。而从统计的角度来讲,只关心数据中出现次数多少,即比例最大的几项,这样的话是没有行为的时序、周期与行为变化特性。而我们所关心的不仅仅是用户常看的节目,更关注用户行为的周期性,与行为的变更性。这样在同一个用户类中的会有很大概率会有相同的行为变更性。当然,我们也可以用其他方式来构建行为模式结构数据,只要所述行为模式结构数据包括行为类型即可。

[0069] 需要注意的是,这里,所述结构性数据(即结构数据)是指由 $P(c,s,n,m,k)$ 结构用户数据分析得出的用户行为变更结构,得到的数据集为 (a_1,a_2,\dots,a_n) ,其定义与线性表相同, n 个元素之间存在序偶关系。其中 a 为行为类别,这样我们得到的数据之间是有序先后关系的,其中就包含了用户行为变化特征。当然我们在实现的时候即可用线性存储方式,也可以使用链式存储方式。

[0070] 在第二步骤S2中,将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户数据集。聚类是将数据分类到不同的类的一个过程,使同一个类中的对象有较大的相似性,不同类的个体有较大的差异性。在本发明第一实施例中,我们将存储在用户行为模式数据库中的用户行为模式结构数据中包含的行为类型(即c)进行聚类分析。其具体设定可以为:行用户行为模式数据库中用于聚类的数据源 $D1=(U,P)$ 其中, $U=\{\text{用户1、用户2} \dots \text{用户n}\}$ 代表用户集合;则行为类型进行聚类的相似度表达式为 $\text{sim}(i,j)=\frac{C_i \cap C_j}{C_i \cup C_j}$; 其中, C_i, C_j 表示U中第i、j个用户的行为

模式类集合。总之,第一次聚类分析是对结构数据 $D1=(U,P), P(c,s,n,m,k)$ 的行为类型c按所述的用户行为类型相似度表达式进行聚类分析,从而生成按类型相似的聚类用户数据集。

[0071] 然后,对按类型相似的聚类用户数据进行第二次按变化结构特征聚类,并生成具有相似行为变迁的用户簇集数据集,即第三步骤S3。在本发明第一实施例中,其具体在于:定义所述按类型用户行为数据集为 $D2=(U,S)$,其中 $U=\{\text{用户1、用户2} \dots \text{用户n}\}$; $S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$, s_{ij} 为第i个用户的第j个行为模式起点, n_{ij} 第i个用户的第j个行为为模式周期, m_{ij} 第i个用户的第j个行为为周期数;定义则聚类的变化特征相似度表达式为

$\text{sim}(i,j)=\frac{\sum w_{ij}}{|C_i \cup C_j|}$, 其中, w_{ij} 的值为1时代表用户i与j用户行为模式类型相同且次序相

同,否则为0; $|C_i \cup C_j|$ 为用户i与用户j存在的类型元素总个数。例如对于(A->B->C->D)与(A->B->D) $\text{sim}(i,j)=3/4$,因为总数为4的节点相似节点有3个。总之,所述第二次聚类即对结构数据 $D2=(U,S), S=\{(s_{ij}, s_{ij}+n_{ij} \times m_{ij})\}$,按所述的变化特征相似度表达式进行聚类分析。

[0072] 最后,所述步骤为S4,输出用户聚类簇集结果数据。从而可以通过所输出的数据中蕴含着的单纯统计数据中没有的用户行为变迁信息,使得对用户的描述会更加的完备,而且可以很方便的进行用户之间的协同分析,通过近似用户的变化来预测用户的可能变化。

[0073] 进一步地,我们可以在所述第一次、第二次聚类分析可采用最小生成树聚类方法(MST)。其具体可以采用如下步骤进行:

[0074] A、用相似度表达式,算出所有点之间的权值;

[0075] B、使用MST聚类构造算法,生成最小生成树;

[0076] C、设定可容忍差异权值,大与或等于其值的边定义为边界边;

[0077] D、打断所有边界边;根据打断数量来确定最终生成的聚类数。

[0078] 其中,所述步骤B中基于MST聚类的最小生成树构造算法可以为普里姆(Prim)算法(时间复杂度为 $O(N^2)$)、克鲁斯卡尔(Kruskal)算法(时间复杂度为 $O(e \log e)$ e为图边数)等。

[0079] 下面以一个具体例子来描述上述基于MST的聚类算法。下面给出了一个关于7个点的差异矩阵,如图3所示:7个点分别代表7个用户,

[0080] 在所述差异度矩阵中,第i行第j列的数字代表的是用户i与用户j之间的差异度。如:第2行第3列的数字0.27表示第二个用户和第三个用户的差异度为0.27。

[0081] 为了说明问题本例,将边的数量限定在偶数(这并不会影响最终结果),图4(a)为

由向量差异矩阵形成的无向图,图中X1到X7分别表示第一个到第七个用户。图4(b)为图4(a)的最小生成树,图4(c)是“可容忍差异”权值大于0.3的聚类结果图,而图4(d)是由“可容忍差异”权值大于0.27的聚类结果图。

[0082] 由聚类结果图可以看出数据被聚类生成聚类特征数据的过程,对数据进行两次MST聚类方案就可以获得具有相同行为变迁的用户聚类簇集。用无向图解决聚类问题的优势在于,从空间角度来想问题更加直观,可理解性较强;另一方面在对新用户数据更新时,只需要对新用户数据进行重新计算,无需计算以前的用户数据,那么进行新用户分类的时候速度会很快。当然使用本算法在计算差异度时也会自身的局限性,数据量在很小的时候,一个点的变化会对其有相当大的影响。

[0083] 这里举例说明,(动作,喜剧,爱情,...)是现有视频节目的基本类型集。在分析中,用户的收看节目数据集包含了这些类型的行为,则用户的行为就分为这类型集。比如:A用户在一段时间内看的电影只有{喜剧,爱情}那这个用户的类型集就是:{喜剧,爱情},B用户在一段时间内看的电影只有{动作,恐怖}那这个用户的类型集就是:{动作,恐怖},同理C用户{喜剧,动作、爱情},D用户{动画,科幻},...,这里A用户与C用户就是属于一个基本类型集,也就是所述的按行为模式中所包含的行为类型特征进行第一次聚类,会将用户A与C聚到同一簇集之中,因为他们两个是相似度较高。

[0084] 进一步地,对具有相同基本类型集的用户,再从观看时序上计算其相似性将相似度较高的用户聚为一类,同一类中的用户之间历史行为变迁度相似度较高,那么一个用户的下一次行为就很可能是其它用户下一次的行,例如用户一与用户二是在同一类中,两者的观看历史序列为:用户一{动作,爱情,喜剧},用户二{动作,爱情},因为两者在同一类中,所以用户一的喜剧的行为很可能是用户二下一次的行,当然这只是一个简单的例子,有可能这个用户观看的记录之间的联系很复杂。

[0085] 比如,某用户喜欢赵本山,那么会分析出一个描述集,这个描述集表示的是这个新的描述所具备的元素有那些,比如我们定义一个新的描述为赵家军={赵本山,小沈阳,二人转},意思是该用户喜欢赵家军这一类型的,而这该类型的要素是:赵本山,小沈阳,二人转。

[0086] 本方法引用了相似点之和描述用户收看节目的程度,如user1(A->B->C->D)与user2(A->B->D)这样相似点之和为3,其中ABCD代表用户观看的类型。如果user2(B->A->D)因为B不满足序列要求不是在A后,所以相似点之和为2。本方法按照该相似点之和对数据集进行进一步的聚类,也就是按类型用户行为数据进行的第二次聚类。

[0087] 另外,本发明还提供了一种用户行为分析预测方法,如图5所示,所述方法包括以下步骤:

[0088] A1、建立用户行为分析预测系统,所述用户行为分析预测系统,存储有按变化结构特征用户行为数据集;所述的按变化结构特征用户行为数据集,采用如是述的用户行为分析方法生成,这里就不再赘述了。

[0089] B1、收集当前用户行为数据;

[0090] C1、将所述用户行为数据输入所述的用户行为分析预测系统,所述的用户行为分析预测系统生成按变化特征用户行为进行聚类的数据聚类簇集,并通过聚类结果中其它用户预测用户行为变迁的可能性。

[0091] 在所述用户行为分析预测系统中,通过存储有按变化结构特征用户行为数据集,对输入的即时用户行为数据进行分析,通过近似用户的变化来预测输入的即时用户行为数据的用户的可能变化,即预测用户行为变迁的可能性。

[0092] 在具体应用时,我们可以将上述用户行为分析预测方法应用到电视节目推广技术领域,为用户提供人性化的消费品和服务。

[0093] 请继续参阅图6,其为本发明的智能电视节目推送系统的结构框图,如图所示,所述的系统包括:智能电视终端100、用户行为聚类分析引擎200和节目推送云端服务器300。所述智能电视终端100、用户行为聚类分析引擎200和节目推送云端服务器300依次连接。

[0094] 其中,所述智能电视终端100用于播放节目与采集即时用户行为数据;在本实施例中,所述智能电视终端100包括用户即时行为数据采集与传输装置,用于与所述的用户行为聚类分析引擎建立网络连接,采集与发送用户即时行为数据。例如:可以通过图像采集装置或观看统计装置来作为用户即时行为数据采集,然后通过与其相连的传输装置(无线、有线等等)与所述的用户行为聚类分析引擎建立网络连接,采集与发送用户即时行为数据。

[0095] 所述用户行为聚类分析引擎200用于根据所述的即时用户行为数据进行包括按类型、按变化结构特征的两次聚类分析,将具有相同历史行为变迁的用户分为一类形成簇集,作为后期推荐数据集,其具体方法可以参见上述的用户行为分析方法。在本实施例中,所述的用户行为聚类分析引擎200包括:数据接收装置和数据聚类分析装置,所述数据接收装置用于与所述的智能电视终端建立网络连接,接收与存储所述的发送用户即时行为数据;所述数据聚类分析装置用于根据用户行为分析方法对所述的用户即时行为数据进行聚类分析,将用户进行分类,具有相似变迁行为的用户聚为一类。

[0096] 所述节目推送云端服务器300用于根据聚类结果集推送用户潜在兴趣节目。在本实施例中,所述节目推送云端服务器300包括用户潜在兴趣节目挖掘装置和节目推送装置,所述的用户潜在兴趣节目挖掘装置与所述的用户行为聚类分析引擎建立网络连接,并采用上述的用户行为分析预测方法,根据挖掘分析出用户聚类集,并输出挖掘结果数据集;节目推送装置,用于与所述的智能电视终端建立网络连接,直接推送用户潜在兴趣节目或再经过协同过滤(也称协同分析)对潜在兴趣节目进行排序然后再进行推送。

[0097] 进一步地,所述节目推送装置经过协同过滤对潜在兴趣节目进行排序然后再进行推送的具体步骤如下:

[0098] A2、列举出潜在兴趣节目;

[0099] B2、根据节目推荐度对潜在兴趣节目进行排序;

[0100] C2、将排序后的潜在兴趣节目推荐给用户。

[0101] 其中,所述步骤B中节目推荐度P的计算方法如下:

$$[0102] \quad P = \frac{\sum_{j=1}^n Sim_{ij} * I_j}{\sum_{j=1}^n Sim_{ij}} ;$$

[0103] 所述P的取值范围在0到1之间;j表示用户数量,其取值范围在1到n之间; I_j 表示用户j是否观看过,当观看时 I_j 为1,否则为0; Sim_{ij} 表示目标用户i与比对对象j之间的相似度。

这种推荐方法在上述二次聚类的结果上进行的,从而大大减少了算法的运算量(现有技术的推荐度的算法有多种,其算法与本申请相比复杂程度高了很多),且保证了向用户推荐节目的准确性,为用户带来了更好的服务。

[0104] 综上所述,本发明提供的用户行为分析方法、分析预测方法及电视节目推送系统,首先,对历史用户行为数据进行分析,提取出用户行为模式结构数据,并存储在用户行为模式数据库中;然后,将存储在用户行为模式数据库中的用户行为模式结构数据按行为类型特征进行第一次聚类,并生成按类型相似的聚类用户数据集;再对所述的按类型相似的聚类用户数据进行第二次按变化结构特征进行聚类,并生成具有相似行为变迁的用户簇集数据集;最后输出用户聚类结果数据。因为第二次聚类考虑了用户行为的变迁时序特征,其中蕴含着单纯统计数据中没有的用户行为变迁信息,使得对用户的描述会更加的完备,而且最后得到的聚类结果集可以很方便的进行用户之间的协同分析,通过近似用户的变化来预测用户的可能变化,并在电视节目推广领域,为用户推送潜在兴趣节目。

[0105] 需要说明的是,本用户行为分析方法基于变化结构的数据分析方法,本分析预测方法基于所述用户行为分析方法,本电视节目推送系统基于所述分析预测方法。

[0106] 可以理解的是,对本领域普通技术人员来说,可以根据本发明的技术方案及其发明构思加以等同替换或改变,而所有这些改变或替换都应属于本发明所附的权利要求的保护范围。

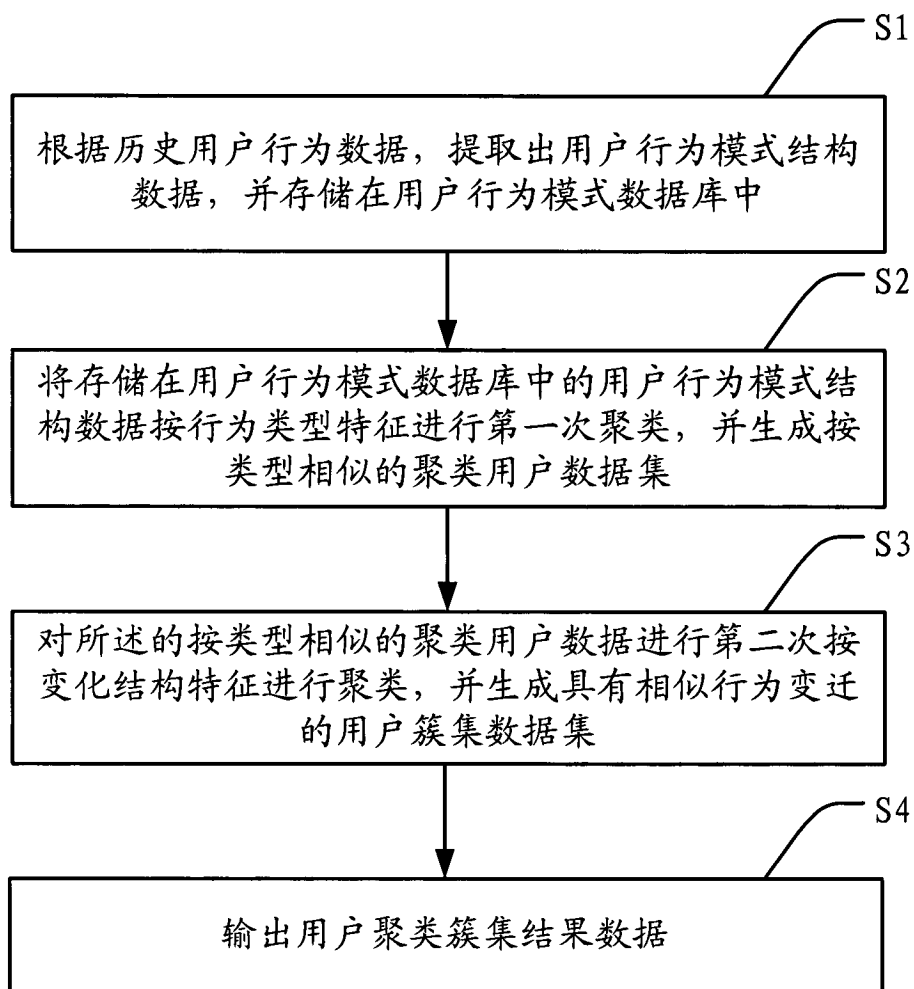


图1

P(c,s,n,m,k)结构		0.0	0.03	0.30	0.60	0.62	0.67	0.70
类别数据		0.03	0.0	0.27	0.43	0.47	0.50	0.53
行为起点		0.30	0.27	0.0	0.57	0.33	0.37	0.40
行为间隔		0.60	0.43	0.57	0.0	0.17	0.20	0.23
行为周期		0.62	0.47	0.33	0.17	0.0	0.10	0.13
行为时间长度		0.67	0.50	0.37	0.2	0.10	0.0	0.07
		0.70	0.53	0.40	0.23	0.13	0.07	0.0

图3

图2

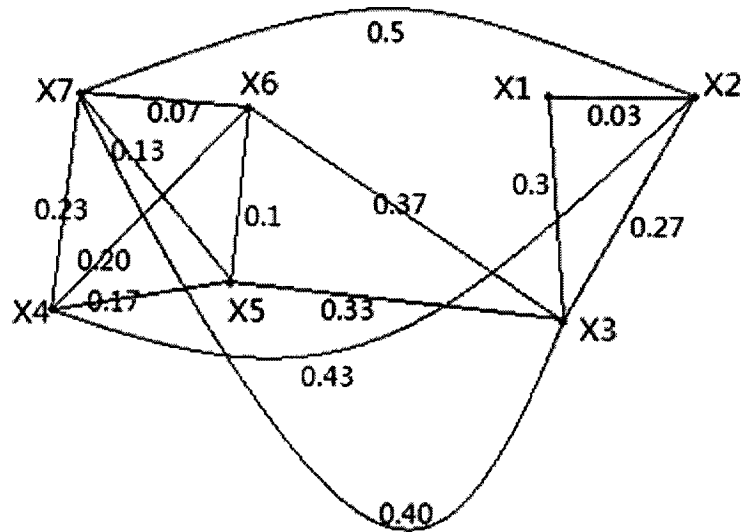


图4(a)

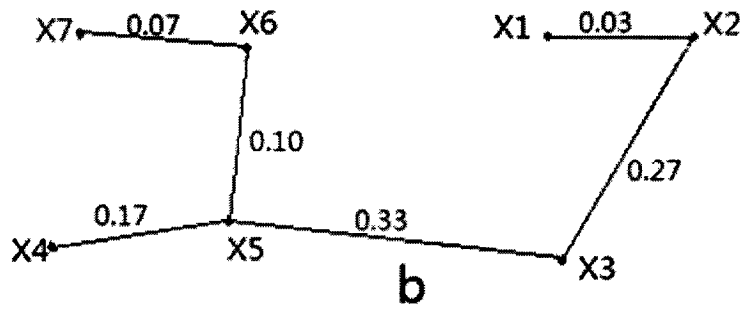


图4(b)

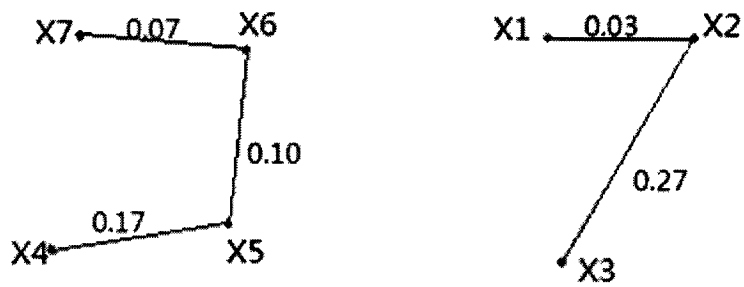


图4(c)

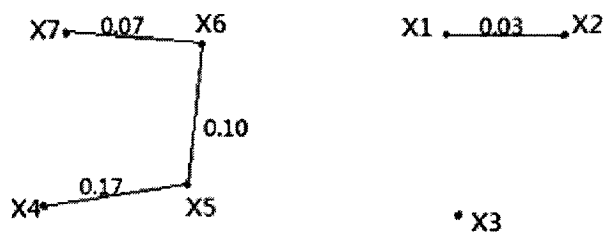


图4(d)

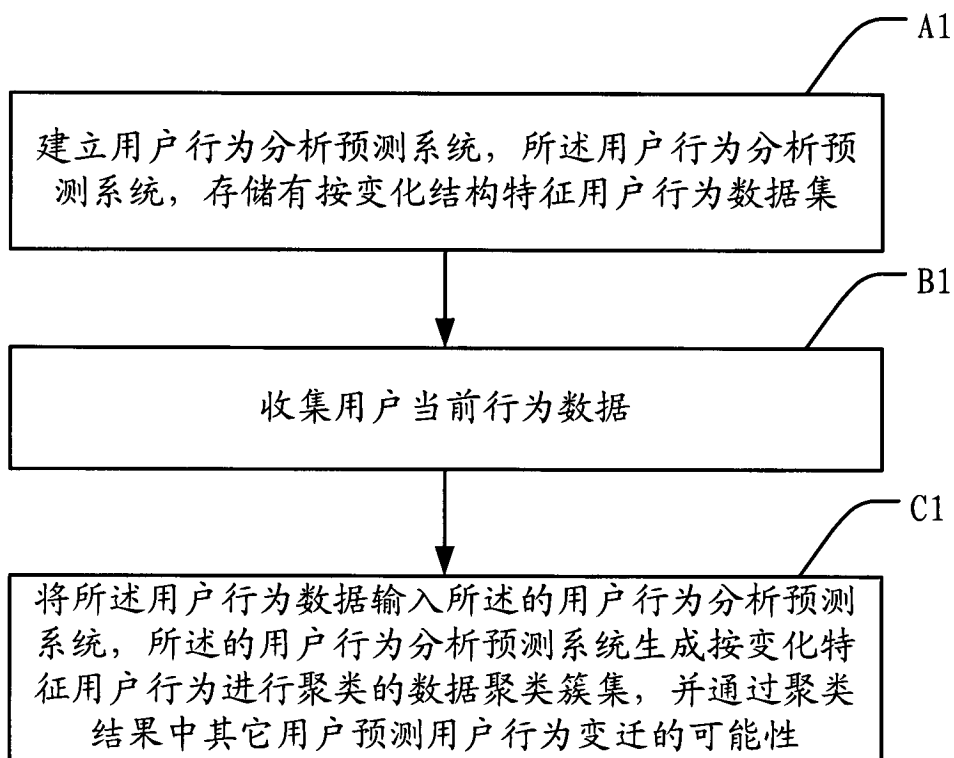


图5

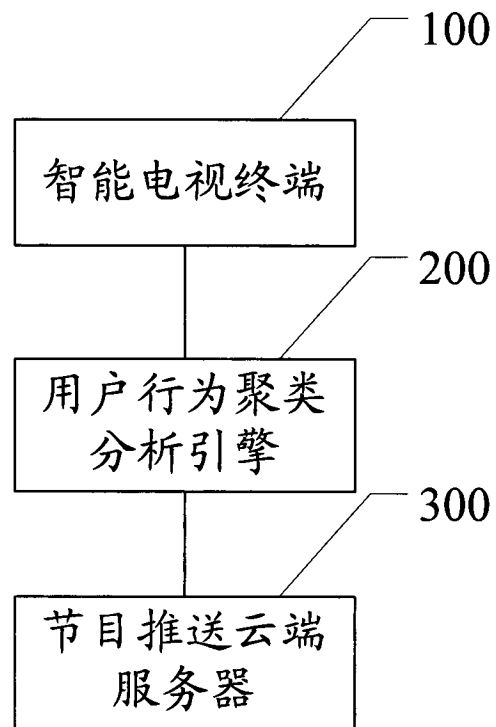


图6