(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107609786 A
(43)申请公布日 2018.01.19

(21)申请号 201710883112.X

(22)申请日 2017.09.26

(71)申请人 宿州学院
地址 234000 安徽省宿州市汴河中路49号（西区）

(72)发明人 汪材印 崔琳 梁楠楠 谈成访 潘正高 彭智超 戚溪溪

(74)专利代理机构 西安铭泽知识产权代理事务所(普通合伙) 61223
代理人 李振瑞

(51)Int.Cl.
G06Q 10/06(2012.01)
G06Q 30/02(2012.01)

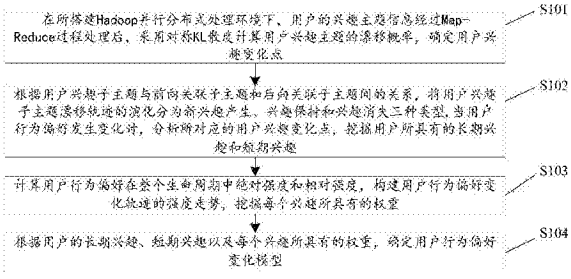
权利要求书2页 说明书5页 附图4页

(54)发明名称

一种在线社会网络下构建用户行为偏好变化模型的方法

(57)摘要

本发明公开了一种在线社会网络下构建用户行为偏好变化模型的方法。该方法包括：在所搭建的Hadoop并行分布式处理环境下，用户的兴趣主题信息经过Map-Reduce过程处理后，采用对称KL散度计算用户兴趣主题的漂移概率，确定用户兴趣变化点；挖掘用户所具有的长期兴趣和短期兴趣；挖掘每个兴趣所具有的权重；根据用户的长期兴趣、短期兴趣以及每个兴趣所具有的权重，确定用户行为偏好变化模型。本发明研究用户在线社会网络下的行为偏好变化，了解用户在不同时空情景下的个性化偏好，并根据当前时刻的情景信息和用户自身兴趣，从浩瀚的在线社会网络环境中发现用户真正感兴趣的信息资源，满足用户对信息的个性化需求，切实地提升用户体验。



CN 107609786 A

1.一种在线社会网络下构建用户行为偏好变化模型的方法,其特征在于,包括:

在所搭建Hadoop并行分布式处理环境下,用户的兴趣主题信息经过Map-Reduce过程处理后,采用对称KL散度计算用户兴趣主题的漂移概率,确定用户兴趣变化点;

根据用户兴趣子主题与前向关联子主题和后向关联子主题间的关系,将用户兴趣子主题漂移轨迹的演化分为新兴趣产生、兴趣保持和兴趣消失三种类型,当用户行为偏好发生变化时,分析所对应的用户兴趣变化点,挖掘用户所具有的长期兴趣和短期兴趣;

计算用户行为偏好在整个生命周期中绝对强度和相对强度,构建用户行为偏好变化轨迹的强度走势,挖掘每个兴趣所具有的权重;

根据用户的长期兴趣、短期兴趣以及每个兴趣所具有的权重,确定用户行为偏好变化模型。

2.如权利要求1所述的在线社会网络下构建用户行为偏好变化模型的方法,其特征在于,所述采用对称KL散度计算用户兴趣主题的漂移概率,判断用户兴趣变化点,包括:

用户兴趣通过使用用户情景日志中特征词汇的出现概率表示,判断不同时期用户兴趣主题间的语义相似性对应于度量两个概率分布间的近似程度;令滑动窗口包含N个时间片,把时间片t中的用户所具有的兴趣j记作T_j^t,原KL散度是不对称的,但不同时期用户兴趣主题间的语义相似性是对称的,即对于任意用户兴趣主题T_j^t和T_l^m, T_j^t和T_l^m的相似性与T_l^m和T_j^t的相似性相等,改进原KL散度,确定基于对称KL散度的用户兴趣子主题相似度为:

$$Sim(T_j^t,T_l^m)=-\frac{1}{2}KL(T_j^t\|T_l^m)-\frac{1}{2}KL(T_l^m\|T_j^t)$$

$$=-\frac{1}{2}\sum_{w\in V}p(w)\log\frac{p(w)}{q(w)}-\frac{1}{2}\sum_{w\in V}q(w)\log\frac{q(w)}{p(w)}$$

其中,p(w)和q(w)分别表示特征词汇w在用户兴趣子主题T_j^t和T_l^m中的出现概率,V代表一个词汇字典的集合。

3.如权利要求2所述的在线社会网络下构建用户行为偏好变化模型的方法,其特征在于,

所述前向关联用户兴趣子主题为:滑窗内各时间片i(i=t-N,…,t-1)中与T_j^t相似度最大的用户兴趣子主题,记作$forward(T_j^t)$;

所述后向关联用户兴趣子主题为:滑窗内各时间片i(i=t+1,…,t+N)中与T_j^t相似度最大的用户兴趣子主题,记作$backward(T_j^t)$。

4.如权利要求3所述的在线社会网络下构建用户行为偏好变化模型的方法,其特征在于,

所述新兴趣产生,包括:对于用户兴趣子主题T_j^t,若不存在前向关联兴趣子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε,既不满足$Sim(T_j^t,T_l^m)\geq\varepsilon$,则$T_j^t$是在时间片t中产生的新兴趣主题;

所述兴趣保持,包括:对于用户兴趣子主题T_j^t,若存在前向关联兴趣子主题

$T_l^m = forward(T_j^t)$使得T_j^t和T_l^m的相似度大于阈值ε，既$Sim(T_j^t, T_l^m) \geq \varepsilon$，并且$T_j^t$也是$T_l^m$的后向关联兴趣子主题，即$T_j^t = backward(T_l^m)$，则$T_j^t$是$T_l^m$的后继，用户的兴趣没有发生太大变化；

所述兴趣消失，包括：对于用户兴趣子主题T_j^t，若不存在后向关联子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε，既不满足$Sim(T_j^t, T_l^m) \geq \varepsilon$，则主题$T_j^t$在时间片t中消亡，用户已不再具有此兴趣。

5.如权利要求1所述的在线社会网络下构建用户行为偏好变化模型的方法，其特征在于，

所述绝对强度包括：设$d_i = \{d_{i1} \cdots d_{iM}\}$表示移动用户情景日志$d_i$的单词序列，M表示用户情景日志$d_i$所包含的单词个数，i表示用户情景日志$d_i$所表达的用户兴趣主题；用户兴趣主题i在时间t的绝对强度采用如下公式：

$$ABIntensity(i,t) = \frac{1}{N} \sum_{t' \in [t-N, t-1]} \sum_{j=1}^{M} \delta(|d_{ij}|_{t'}, i)$$

所述相对强度包括：用户兴趣主题i在时间t的相对强度采用如下公式：

$$REIntensity(i,t) = \frac{ABIntensity(i,t)}{\sum_{p=1}^{K} ABIntensity(p,t)} = \frac{\sum_{t' \in [t-N, t-1]} \sum_{j=1}^{M} \delta(|d_{ij}|_{t'}, i)}{\sum_{t' \in [t-N, t-1]} |d_{t'}|}$$

其中，t′为t-N与t-1之间的任一时间片，K为用户兴趣主题的个数，p为用户兴趣主题K中的其中任意一个主题。

一种在线社会网络下构建用户行为偏好变化模型的方法

技术领域

[0001] 本发明涉及在线社会网络领域，更具体的涉及一种在线社会网络下构建用户行为偏好变化模型的方法。

背景技术

[0002] 用户行为偏好变化建模(即用户兴趣变化建模)指对用户有关兴趣及其行为信息进行归纳，基于此演绎出可描述其行为变化特征的用户模型，它是推荐系统的关键技术之一，只有准确地把握用户兴趣变化，才能设计出高效的推荐算法，从而提升用户体验，用户行为偏好变化建模是进行个性化检索与推荐的前提，随着个性化服务的不断发展和研究的逐步深入，人们开始意识到用户兴趣及其偏好变化描述对个性化服务的重要意义。

[0003] 目前，对在线社会网络下的个性化推荐问题虽有一定程度的研究，但是，由于在在线社会网络下，用户的行为偏好会经常随时间发生变化，而传统建模方法大都采用静态的研究方法，没有考虑在线社会网络下用户行为偏好变化的问题。

[0004] 综上所述，现有技术中的用户兴趣建模，大都采用静态的研究方法，没有考虑在线社会网络下行为偏好变化的问题。

发明内容

[0005] 本发明实施例提供一种在线社会网络下用户行为偏好变化模型构建的方法，用以解决现有技术中采用静态的研究方法，没有考虑用户行为偏好变化的问题。

[0006] 本发明实施例提供一种在线社会网络下构建用户行为偏好变化模型的方法，包括：

[0007] 在所搭建Hadoop并行分布式处理环境下，用户的兴趣主题信息经过Map-Reduce过程处理后，采用对称KL散度计算用户兴趣主题的漂移概率，确定用户兴趣变化点；

[0008] 根据用户兴趣子主题与前向关联子主题和后向关联子主题间的关系，将用户兴趣子主题漂移轨迹的演化分为新兴趣产生、兴趣保持和兴趣消失三种类型；当用户行为偏好发生变化时，分析所对应的用户兴趣变化点，挖掘用户所具有的长期兴趣和短期兴趣；

[0009] 计算用户行为偏好在整个生命周期中绝对强度和相对强度，构建用户行为偏好变化轨迹的强度走势，挖掘每个兴趣所具有的权重；

[0010] 根据用户的长期兴趣、短期兴趣以及每个兴趣所具有的权重，确定用户行为偏好变化模型。

[0011] 较佳地，所述采用对称KL散度计算用户兴趣主题的漂移概率，判断用户兴趣变化点，包括：

[0012] 用户兴趣通过使用用户情景日志中特征词汇的出现概率表示，判断不同时期用户兴趣主题间的语义相似性对应于度量两个概率分布间的近似程度；令滑动窗口包含N个时间片，把时间片t中的用户所具有的兴趣j记作T_j^t，原KL散度是不对称的，但不同时期用户兴趣主题间的语义相似性是对称的，即对于任意用户兴趣主题T_j^t和T_l^m，T_j^t和T_l^m的相似性与

T_l^m和T_j^t的相似性相等，改进原KL散度，确定基于对称KL散度的用户兴趣子主题相似度为：

[0013]
$$Sim(T_j^t,T_l^m)=-\frac{1}{2}KL(T_j^t\|T_l^m)-\frac{1}{2}KL(T_l^m\|T_j^t)$$
$$=-\frac{1}{2}\sum_{w\in V}p(w)\log\frac{p(w)}{q(w)}-\frac{1}{2}\sum_{w\in V}q(w)\log\frac{q(w)}{p(w)}$$

[0014] 其中，p(w)和q(w)分别表示特征词汇w在用户兴趣子主题T_j^t和T_l^m中的出现概率，V代表一个词汇字典的集合。

[0015] 较佳地，

[0016] 所述前向关联子主题为：滑窗内各时间片i(i＝t-N,…,t-1)中与T_j^t相似度最大的用户兴趣子主题，记作$forward(T_j^t)$；

[0017] 所述后向关联子主题为：滑窗内各时间片i(i＝t+1,…,t+N)中与T_j^t相似度最大的用户兴趣子主题，记作$backward(T_j^t)$。

[0018] 较佳地，

[0019] 所述新兴趣产生，包括：对于用户兴趣子主题T_j^t，若不存在前向关联兴趣子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε，既不满足$Sim(T_j^t,T_l^m)\geq\varepsilon$，则$T_j^t$是在时间片t中产生的新兴趣主题；

[0020] 所述兴趣保持，包括：对于用户兴趣子主题T_j^t，若存在前向关联兴趣子主题$T_l^m=forward(T_j^t)$使得T_j^t和T_l^m的相似度大于阈值ε，既$Sim(T_j^t,T_l^m)\geq\varepsilon$，并且$T_j^t$也是$T_l^m$的后向关联兴趣子主题，即$T_j^t=backward(T_l^m)$，则$T_j^t$是$T_l^m$的后继，用户的兴趣没有发生太大变化；

[0021] 所述兴趣消失，包括：对于用户兴趣子主题T_j^t，若不存在后向关联子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε，既不满足$Sim(T_j^t,T_l^m)\geq\varepsilon$，则$T_j^t$是在时间片t中消亡，用户已不再具有此兴趣。

[0022] 较佳地，

[0023] 所述绝对强度包括：设$d_i=\{d_{i1}\cdots d_{iM}\}$表示移动用户情景日志$d_i$的单词序列，M表示用户情景日志$d_i$所包含的单词个数，i表示用户情景日志$d_i$所表达的用户兴趣主题；用户兴趣主题i在时间t的绝对强度采用如下公式：

[0024]
$$ABIntensity(i,t)=\frac{1}{N}\sum_{t'\in[t-N,t-1]}\sum_{j=1}^{M}\delta(|d_{ij}|_{t'},i)$$

[0025] 所述相对强度包括：用户兴趣主题i在时间t的相对强度采用如下公式：

[0026]
$$REIntensity(i,t)=\frac{ABIntensity(i,t)}{\sum_{p=1}^{K}ABIntensity(p,t)}=\frac{\sum_{t'\in[t-N,t-1]}\sum_{j=1}^{M}\delta(|d_{ij}|_{t'},i)}{\sum_{t'\in[t-N,t-1]}|d_{t'}|}$$

[0027] 其中，t′为t-N与t-1之间的任一时间片，K为用户兴趣主题的个数，p为用户兴趣主

题K中的其中任意一个主题。

[0028] 本发明实施例中，提供一种在线社会网络下构建用户行为偏好变化模型的方法，与现有技术相比，其有益效果为：本发明在所搭建的Hadoop并行分布式处理环境下，研究用户的行为偏好变化，了解用户在不同时间下的个性化偏好，并根据当前时刻的情景信息和用户自身兴趣，从浩瀚的在线社会网络环境中发现用户真正感兴趣的信息资源，满足用户对信息的个性化需求，切实地提升用户体验；本发明的研究对提高个性化服务的应用效果和寻求在线社会网络下个性化信息服务技术的突破具有重要价值，以期推动在线社会网络应用和服务向智能化进一步发展。

附图说明

[0029] 图1为本发明实施例提供的一种在线社会网络下构建用户行为偏好变化模型的方法流程图；

[0030] 图2为本发明实施例提供的用户兴趣子主题产生示意图；

[0031] 图3为本发明实施例提供的用户兴趣子主题继承示意图；

[0032] 图4为本发明实施例提供的用户兴趣子主题消亡示意图。

具体实施方式

[0033] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0034] 图1为本发明实施例提供的一种在线社会网络下构建用户行为偏好变化模型的方法流程图。如图1所示，该方法包括：

[0035] 步骤S101，在所搭建的Hadoop并行分布式处理环境下，采用对称KL散度计算用户兴趣主题的漂移概率，确定用户兴趣变化点。

[0036] 在Hadoop集群环境下，用户兴趣的前期预处理被划分为Map过程和Reduce过程。在Map过程中，任务执行时，每一个Map的输入数据来源于HDFS中划分好的一个块Split(每个数据块大小设为64MB)。经过Map过程处理后，Map输出key-value对，设定key对应着用户的不同的兴趣主题T，value是key所对应主题的概率值。随后，在Shuffle阶段，对Map过程的处理结果执行排序与合并操作，然后再把Shuffle处理后的key-value信息输入给Reduce过程进行处理；在Reduce任务执行时，依据26个英文字母的顺序，划分为5个reduce task，按照key值首字母的顺序二次合并结果数据集，并把Map-Reduce操作的最终处理结果输出，以进一步分析用户的行为偏好随着时间和所处情景如何发生变化。

[0037] 在在线社会网络下，尽管用户的行为偏好会随着时间和所处情景发生变化，但不同时期的用户的兴趣主题之间存在一定的关联，这种关联反映为不同时期用户兴趣主题间的语义相似性。因此，用户兴趣使用特征词汇的出现概率表示，判断不同时期用户兴趣主题间的语义相似性对应于度量两个概率分布间的近似程度。令滑动窗口包含N个时间片，把时间片t中的用户所具有的兴趣j记作T_j^t，原KL散度是不对称的，但不同时期用户兴趣主题间的语义相似性是对称的，即对于任意用户兴趣主题T_j^t和T_l^m，T_j^t和T_l^m的相似性与T_l^m和T_j^t的相

似性相等，改进原KL散度，初步构思基于对称KL散度的用户兴趣子主题相似度为：

[0038]
$$Sim(T_j^t,T_l^m)=-\frac{1}{2}KL(T_j^t\|T_l^m)-\frac{1}{2}KL(T_l^m\|T_j^t)$$
$$=-\frac{1}{2}\sum_{w\in V}p(w)\log\frac{p(w)}{q(w)}-\frac{1}{2}\sum_{w\in V}q(w)\log\frac{q(w)}{p(w)} \quad (1)$$

[0039] 其中，p(w)和q(w)分别表示特征词汇w在用户兴趣子主题T_j^t和T_l^m中的出现概率。

[0040] 步骤S102，根据用户兴趣子主题与前向关联子主题和后向关联子主题间的关系，将用户兴趣子主题漂移轨迹的演化分为新兴趣产生、兴趣保持和兴趣消失三种类型，当用户行为偏好发生变化时，分析所对应的用户兴趣变化点，挖掘用户所具有的长期兴趣和短期兴趣。

[0041] 用户兴趣主题具有一定的生命周期，由一组相互关联的用户兴趣子主题组成。令N是时间滑动窗口的大小，对于时间片t中的用户兴趣子主题T_j^t，与时间片t相邻滑窗内用户兴趣子主题间存在关联。滑窗内各时间片i(i＝t-N,…,t-1)中与T_j^t相似度最大的用户兴趣子主题称为T_j^t的前向关联子主题，记作$forward(T_j^t)$；时间片i(i＝t+1,…,t+N)中与T_j^t相似度最大的子主题称为T_j^t的后向关联子主题，记作$backward(T_j^t)$。

[0042] 根据用户兴趣子主题与其前向和后向关联子主题间的关系，将用户兴趣子主题漂移轨迹的演化分为新兴趣产生、兴趣保持和兴趣消失三种类型进行研究，分别介绍如下：

[0043] 用户兴趣子主题产生——新兴趣产生

[0044] 对于用户兴趣子主题T_j^t，若不存在前向关联兴趣子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε，既不满足$Sim(T_j^t,T_l^m)\geq\varepsilon$，则$T_j^t$是在时间片t中产生的新兴趣主题，如图2所示。

[0045] 用户兴趣子主题继承——兴趣保持

[0046] 对于用户兴趣子主题T_j^t，若存在前向关联兴趣子主题$T_l^m=forward(T_j^t)$使得T_j^t和T_l^m的相似度大于阈值ε，既$Sim(T_j^t,T_l^m)\geq\varepsilon$，并且$T_j^t$也是$T_l^m$的后向关联兴趣子主题，即$T_j^t=backward(T_l^m)$，则$T_j^t$是$T_l^m$的后继，如图3所示。这一现象说明用户的兴趣没有发生太大变化。

[0047] 用户兴趣子主题消亡——兴趣消失

[0048] 对于用户兴趣子主题T_j^t，若不存在后向关联子主题T_l^m使得T_j^t和T_l^m的相似度大于阈值ε，既不满足$Sim(T_j^t,T_l^m)\geq\varepsilon$，则$T_j^t$是在时间片t中消亡，如图4所示，说明用户已不再具有此兴趣。

[0049] 步骤S103，计算用户行为偏好在整个生命周期中绝对强度和相对强度，构建用户行为偏好变化轨迹的强度走势，挖掘每个兴趣所具有的权重。

[0050] 步骤S104，根据用户的长期兴趣、短期兴趣以及每个兴趣所具有的权重，确定用户行为偏好变化模型。

[0051] 随着时间的发展，伴随着用户行为偏好发生变化的同时，用户兴趣主题强度也会发生变化。本发明通过计算用户所具有的每个兴趣的绝对强度和相对强度，来挖掘每个兴趣的权重，以判断用户感兴趣的程度是否发生变化。

[0052] 设$d_i=\{d_{i1}\cdots d_{iM}\}$表示移动用户情景日志$d_i$的单词序列，M表示用户情景日志$d_i$所包含的单词个数，i表示用户情景日志$d_i$所表达的用户兴趣主题。用户兴趣主题i在时间t的绝对强度拟采用如下计算方法：

[0053]

$$ABIntensity(i,t)=\frac{1}{N}\sum_{t'\in[t-N,t-1]}\sum_{j=1}^{M}\delta(|d_{ij}|_{t'},i) \quad (2)$$

[0054] 在t′时间片内，当单词$|d_{ij}|_{t'}$属于用户兴趣主题i时，$\delta(|d_{ij}|_{t'},i)=1$，否则$\delta(|d_{ij}|_{t'},i)=0$。

[0055] 用户兴趣主题i在时间t的相对强度拟采用如下计算方法：

[0056]

$$REIntensity(i,t)=\frac{ABIntensity(i,t)}{\sum_{p=1}^{K}ABIntensity(p,t)}=\frac{\sum_{t'\in[t-N,t-1]}\sum_{j=1}^{M}\delta(|d_{ij}|_{t'},i)}{\sum_{t'\in[t-N,t-1]}|d_{t'}|} \quad (3)$$

[0057] 其中，t′为t-N与t-1之间的任一时间片，K为用户兴趣主题的个数，p为用户兴趣主题K中的其中任意一个主题。因此，用户的每一个行为偏好的强度变化可以通过在整个主题生命周期内迭代计算用户兴趣主题i的绝对强度和相对强度进行实现，从而判断每个兴趣的感兴趣的程度是否发生变化。

[0058] 基于以上分析，在挖掘用户的长期兴趣和短期兴趣以及每个兴趣的权重基础之上，就可以构建用户行为偏好变化模型。

[0059] 综上所述，本发明以研究用户在线社会网络下的用户行为偏好变化，了解用户在不同时空情景下的个性化偏好，并根据当前时刻的情景信息和用户自身兴趣，从浩瀚的社会网络环境中发现用户真正感兴趣的信息资源，满足用户对信息的个性化需求，切实地提升用户体验；本发明的研究对提高个性化服务的应用效果和寻求在线社会网络下个性化信息服务技术的突破具有重要价值，以期推动在线社会应用和服务向智能化进一步发展。

[0060] 以上公开的仅为本发明的几个具体实施例，本领域的技术人员可以对本发明进行各种改动和变型而不脱离本发明的精神和范围，倘若本发明的这些修改和变型属于本发明权利要求及其等同技术的范围之内，则本发明也意图包含这些改动和变型在内。

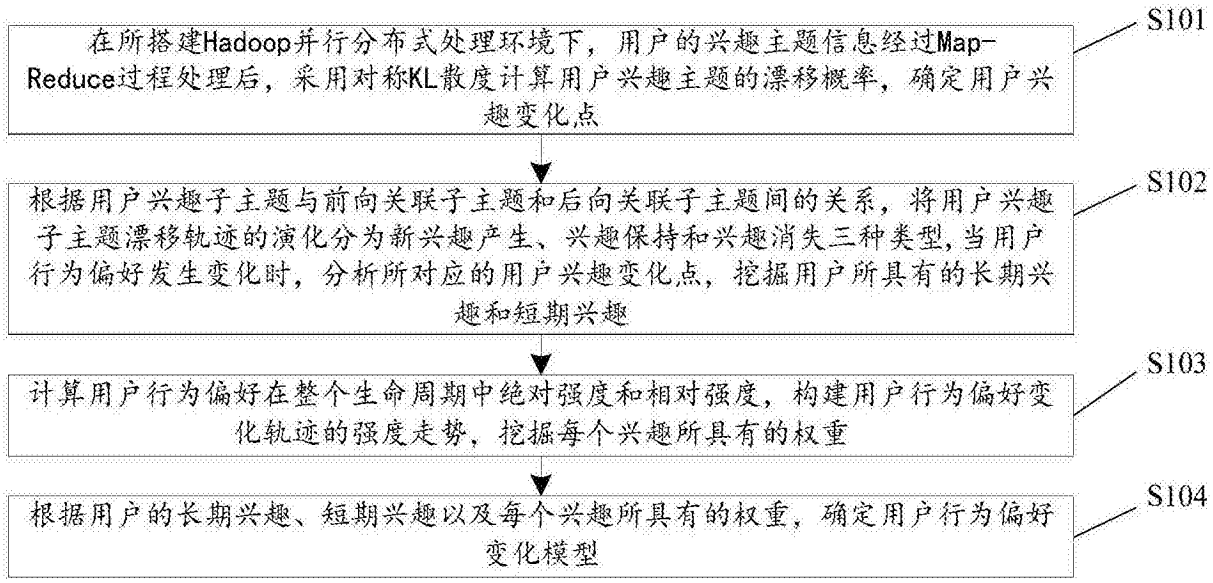

在所搭建Hadoop并行分布式处理环境下，用户的兴趣主题信息经过Map-Reduce过程处理后，采用对称KL散度计算用户兴趣主题的漂移概率，确定用户兴趣变化点
S101
根据用户兴趣子主题与前向关联子主题和后向关联子主题间的关系，将用户兴趣子主题漂移轨迹的演化分为新兴趣产生、兴趣保持和兴趣消失三种类型，当用户行为偏好发生变化时，分析所对应的用户兴趣变化点，挖掘用户所具有的长期兴趣和短期兴趣
S102
计算用户行为偏好在整个生命周期中绝对强度和相对强度，构建用户行为偏好变化轨迹的强度走势，挖掘每个兴趣所具有的权重
S103
根据用户的长期兴趣、短期兴趣以及每个兴趣所具有的权重，确定用户行为偏好变化模型
S104

图1

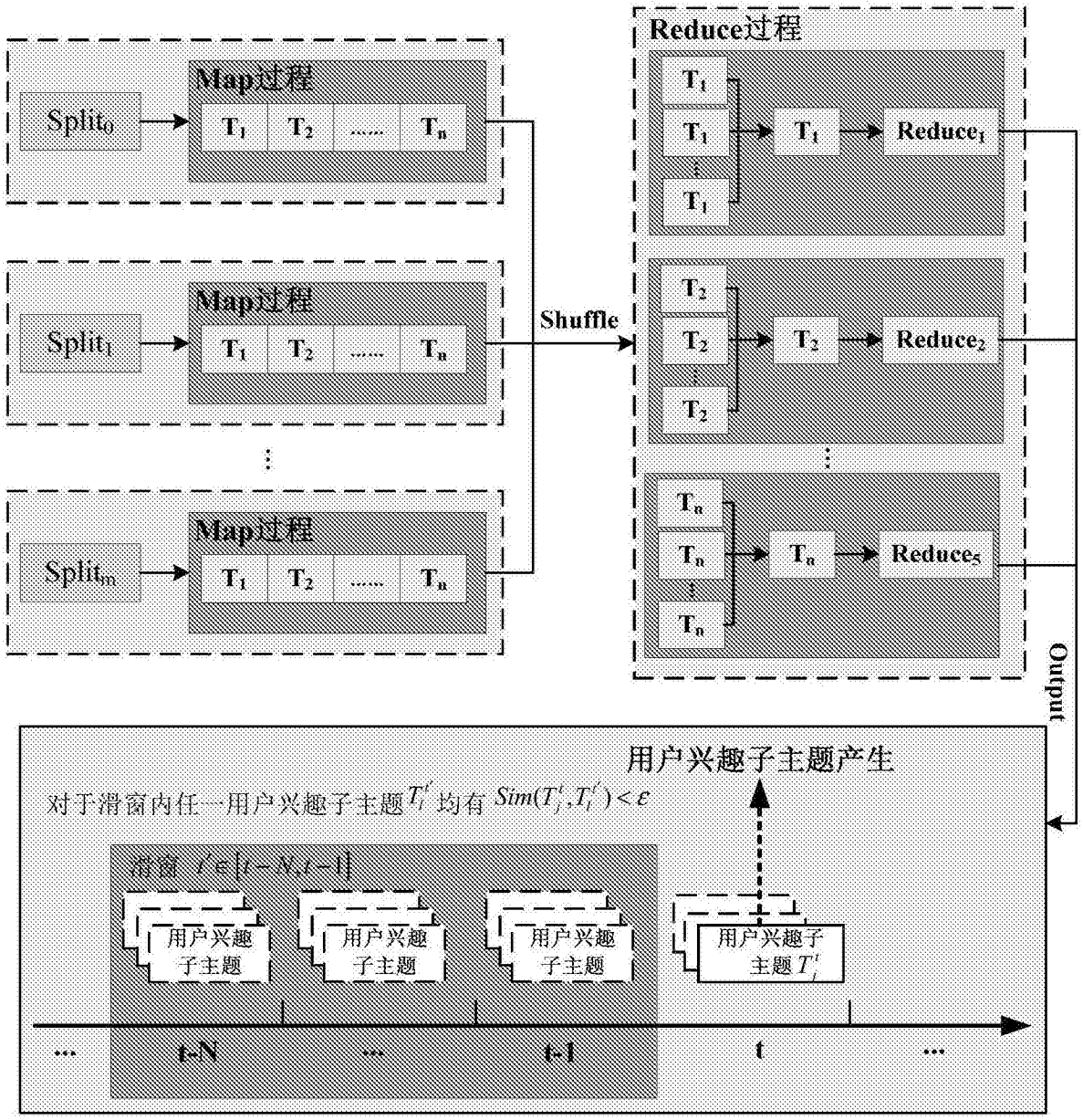

Map过程
Split0
T1 T2 ..... Tn
Split1
Splitm
Shuffle
Reduce过程
T1 T1 T1 Reduce1
T2 T2 T2 Reduce2
Tn Tn Tn Reduce5
Output
用户兴趣子主题产生
对于滑窗内任一用户兴趣子主题 $T_i^{t'}$ 均有 $Sim(T_j^t, T_i^{t'}) < \varepsilon$
滑窗 $t' \in [t-N, t-1]$
用户兴趣子主题
用户兴趣子主题 T_j^t
... t-N ... t-1 t ...

图2

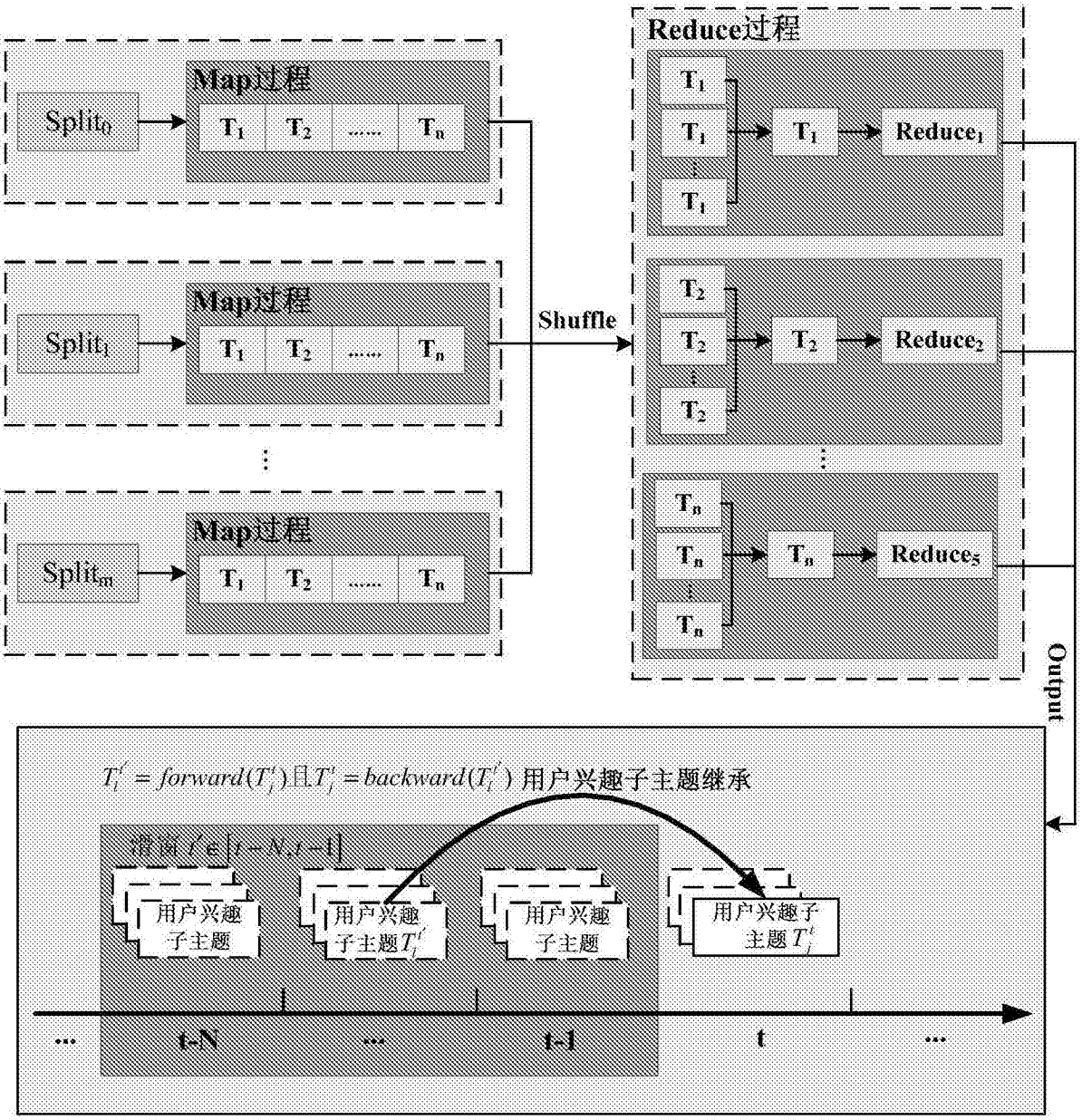

Map过程
Split0
T1 T2 ..... Tn
Splitj
Splitm
Shuffle
Reduce过程
T1 T1 T1 Reduce1
T2 T2 T2 Reduce2
Tn Tn Tn Reduce5
Output
$T_i^{t'} = forward(T_j^t)$且$T_j^t = backward(T_i^{t'})$ 用户兴趣子主题继承
滑窗 $t' \in [t-N, t-1]$
用户兴趣子主题
用户兴趣子主题$T_j^{t'}$
用户兴趣子主题
用户兴趣子主题T_j^t
... t-N ... t-1 t ...

图3

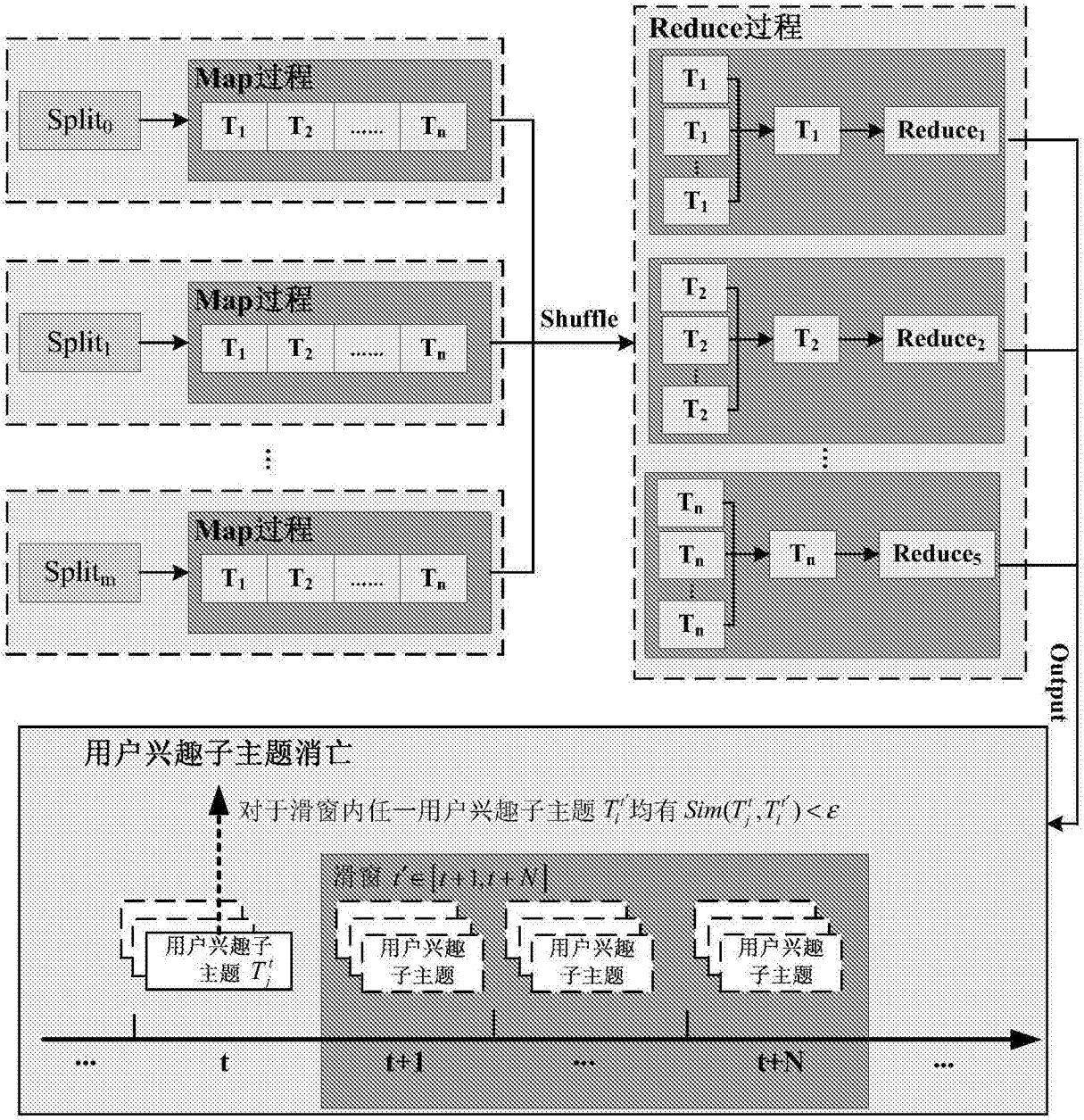

Map过程
Split0
T1 T2 ..... Tn
Split1
Split m
Shuffle
Reduce过程
T1
T2
Tn
Reduce1
Reduce2
Reduce5
Output
用户兴趣子主题消亡
对于滑窗内任一用户兴趣子主题 T_i^t 均有 $Sim(T_j^t, T_i^t) < \varepsilon$
滑窗 $t' \in [t+1, t+N]$
用户兴趣子主题 T_i^t
用户兴趣子主题
t
t+1
t+N

图4