



(12)发明专利

(10)授权公告号 CN 103473291 B

(45)授权公告日 2017.01.18

(21)申请号 201310392446.9

(22)申请日 2013.09.02

(65)同一申请的已公布的文献号

申请公布号 CN 103473291 A

(43)申请公布日 2013.12.25

(73)专利权人 中国科学院软件研究所

地址 100190 北京市海淀区中关村南四街4号

(72)发明人 彭启民 胡堰 胡晓惠

(74)专利代理机构 北京科迪生专利代理有限公司 11251

代理人 杨学明

(51)Int.Cl.

G06F 17/30(2006.01)

(56)对比文件

CN 101414296 A, 2009.04.22,

CN 102819575 A, 2012.12.12,

US 8145636 B1, 2012.03.27,

US 2008250026 A1, 2008.10.09,

审查员 王晓霞

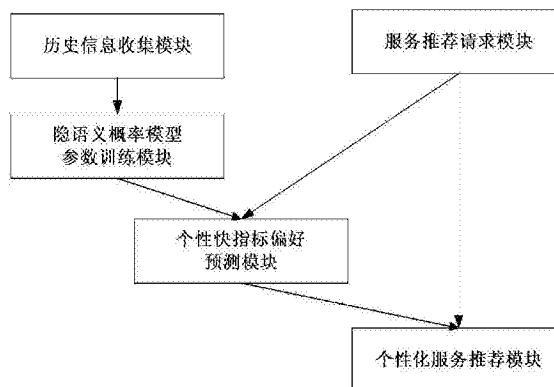
权利要求书4页 说明书7页 附图4页

(54)发明名称

一种基于隐语义概率模型的个性化服务推荐系统及方法

(57)摘要

本发明涉及一种基于隐语义概率模型的个性化服务推荐系统及方法,属于服务计算技术领域,包括以下步骤:确定用于评价一系列功能相似服务的性能优劣的QoS指标体系;建立用户、用户指标偏好以及服务情境三者之间的隐语义概率模型;收集系统内不同用户在不同的服务情境下使用不同功能的服务时通过与系统交互提供的指标偏好信息,并存入历史经验数据库;用已收集的数据训练隐语义概率模型的参数;当用户不熟悉某种服务情境,但需要在该服务情景下调用特定功能服务时,利用已训练的隐语义概率模型对该用户的指标偏好进行预测;根据预测出的用户个性化的QoS指标偏好,对候选服务进行综合筛选,从中选出最贴近该用户需求的服务,从而实现个性化服务推荐。



1. 一种基于隐语义概率模型的个性化服务推荐系统,其特征在于包括:

历史信息收集模块:收集不同用户在不同服务情境下通过层次分析法自主提供的指标偏好比较矩阵,检查该矩阵满足一致性约束后,计算该矩阵最大特征值对应的特征向量,归一化后得出不同用户在不同服务情境下的指标偏好,并将用户、服务情境、指标偏好以三元组形式存入数据库,为隐语义概率模型参数训练模块提供历史训练数据;

隐语义概率模型参数训练模块:从数据库中获取历史信息收集模块收集的用户、服务情境、指标偏好三元组作为训练样本,根据事先设定的隐语义概率模型参数初值,利用EM算法在训练数据上迭代训练模型参数,直至模型参数收敛,保存所得训练参数,提供给个性化指标偏好预测模块;

用户服务推荐请求模块:该模块获取用户登录时提供的个性资料信息及用户登录后录入的需要进行服务推荐的服务情境信息,最后将这两项信息提供给个性化指标偏好预测模块;

个性化指标偏好预测模块:获取用户服务推荐请求模块提供的用户个人信息和服务情境信息,结合隐语义概率模型参数训练模块提供的已经训练好的模型参数,预测该用户在该服务情境下的指标偏好,并将预测的指标偏好提供给个性化服务推荐模块;

个性化服务推荐模块:接收来自个性化指标偏好模块的用户指标偏好,及来自用户服务推荐请求模块的服务情境信息,获取适用于该服务情境的所有功能相似的服务,及服务各维度的QoS指标值,与用户个性化的指标偏好作点积,给出服务的综合排序,作为推荐列表返回给用户,用户根据推荐的结果选择自己感兴趣的项目进行关注。

2. 根据权利要求1所述的基于隐语义概率模型的个性化服务推荐系统,其特征在于:所述隐语义概率模型参数训练模块实现过程:

(1)从数据库中获取历史信息收集模块存入的用户、服务情境、指标偏好三元组,作为历史训练数据;

(2)人工设置用户隐类和服务情境隐类的个数以及模型的参数初值,包括用户隐类和服务情境隐类的先验概率,单个用户在给定不同用户隐类时的条件概率,单则服务情境在给定不同服务情境隐类时的条件概率,以及各个指标偏好在给定不同用户隐类和不同服务情境隐类时的正态分布均值和方差;

(3)用EM算法,结合历史训练数据,对模型参数进行迭代训练,直至收敛,保存模型参数,提供给个性化指标偏好预测模块。

3. 一种基于隐语义概率模型的个性化服务推荐方法,其特征在于实现步骤如下:

步骤1、确定评价服务性能优劣的服务QoS指标体系

所述的服务QoS指标体系是指整个Web服务系统统一采用的用于评价一系列功能相似服务性能优劣所用QoS指标的集合,不同的系统可以根据需要选取适当的QoS指标组成自己的QoS指标体系用于评价服务的性能优劣;

步骤2、建立用户、用户指标偏好以及服务情境三者之间的隐语义概率模型

所述的用户指标偏好是指用户对步骤1中所述服务QoS指标体系中各个QoS指标的偏好程度,对每个QoS指标的偏好程度值介于0到1之间,并且对各个QoS指标的偏好值总和为1;所述的服务情境是指在何种场景下使用何种功能的服务,每一个服务情境 e 用三元组 (w_1, w_2, w_3) 表示,其中 w_1 表示服务完成的与业务无关的基本功能,包括视频功能、导航功能, w_2 表

示用户使用该功能的服务完成的具体业务活动, w_3 表示用户调用服务的终端设备;

所述的用户、用户指标偏好、服务情境三者之间的隐语义概率模型是指单个用户以不同的概率依赖不同的用户隐类而存在, 单则服务情境以不同的概率依赖不同的服务情境隐类存在, 指标偏好同时以不同的概率同时依赖不同的用户隐类和服务情境隐类存在的概率模型, 所述的用户隐类是指非人为事先确定的而是从历史经验数据中学习得到的用户聚簇; 所述的服务情境隐类是指非人为事先确定的而是从历史经验数据中学习得到的服务情境聚簇;

步骤3、收集不同用户在不同的服务情境下使用不同功能服务时自主提供的指标偏好信息, 作为历史经验数据存入数据库, 为训练步骤2中建立的隐语义概率模型的参数作准备, 存储格式为用户、服务情境、指标偏好三元组;

步骤4、用EM算法及已收集的历史经验数据训练隐语义概率模型的参数

所述的隐语义概率模型的参数是指所有用户隐类 $\{U_1, U_2, \dots, U_I\}$ 的先验概率 $P(U_i) (1 \leq i \leq I)$ 、服务情境隐类 $\{E_1, E_2, \dots, E_J\}$ 先验概率 $P(E_j) (1 \leq j \leq J)$ 、给定一个用户隐类 U_i 的情况下单个用户 u 出现的条件概率 $P(u | U_i) (1 \leq i \leq I)$ 、给定一个服务情境隐类 E_j 的情况下单则服务情境 e 出现的条件概率 $P(e | E_j) (1 \leq j \leq J)$ 、给定用户隐类 U_i 和服务情境隐类 E_j 的情况下用户指标偏好向量 r 出现的概率 $P(r | U_i, E_j) (1 \leq i \leq I, 1 \leq j \leq J)$; 其中 I, J 分别表示用户隐类和服务情境隐类的总个数, i, j 分别表示用户隐类和服务情境隐类的编号;

步骤5、获得用户需要服务推荐的请求, 包括个人信息、服务推荐所依赖的服务情境;

步骤6、用已训练的隐语义概率模型预测指定用户在特定服务情境下的未知指标偏好;

步骤7、根据预测出的用户个性化的QoS指标偏好, 对候选服务进行综合筛选, 从而选出最贴近该用户需求的服务, 将推荐结果返回给用户。

4. 根据权利要求3所述的基于隐语义概率模型的个性化服务推荐方法, 其特征在于: 所述步骤3中收集历史经验数据的过程如下:

(1) 用户在其熟悉的服务情境下调用熟悉的web服务时, 直接与系统交互, 使用层次分析法给出自己对各个指标偏好的两两比较结果, 建立比较矩阵;

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix},$$

(2) 验证比较矩阵的一致性是否在可接受的范围内, 若可接受, 进入步骤(3), 否则返回步骤(1);

(3) 计算比较矩阵的最大特征值对应的特征向量, 将该特征向量归一化后所得到的就是用户对各个QoS指标的偏好向量 W_{QoS} , 将用户、服务情境、指标偏好三元组存入数据库。

5. 根据权利要求3所述的基于隐语义概率模型的个性化服务推荐方法, 其特征在于: 所述步骤4中隐语义概率模型的参数训练过程如下,

(1) 从数据库中取出所有的用户、服务情境、指标偏好三元组作为训练数据, 是用户在其熟悉的服务情境下利用层次分析法给出的个性化指标偏好权值, 每个三元组 (u, e, r) 表示用户 u 在服务情境 e 下对服务的各个QoS指标的偏好权重为 r , r 是一个 K 维向量 $(r_1, r_2, \dots,$

r_k),为简单起见,假设权重向量 r 的每一维相互独立并且服从正态分布,则联合概率 $P(u, e, r)$ 用下面公式表示,

$$P(u, e, r) = \sum_i \sum_j P(\mathcal{U}_i) P(\mathcal{F}_j) P(u | \mathcal{U}_i) P(e | \mathcal{F}_j) P(r | \mathcal{U}_i, \mathcal{F}_j),$$

其中 $P(r | \mathcal{U}_i, \mathcal{F}_j)$ 为 K 维正态分布, i, j 分别表示用户隐类和服务情境隐类编号, K 是QoS指标体系中指标个数;

(2)E步计算隐类的联合后验概率,

$$P(\mathcal{U}_p, \mathcal{F}_q | u, e, r) = \frac{[P(\mathcal{U}_p) P(\mathcal{F}_q) P(u | \mathcal{U}_p) P(e | \mathcal{F}_q) P(r | \mathcal{U}_p, \mathcal{F}_q)]^b}{\sum_i \sum_j [P(\mathcal{U}_i) P(\mathcal{F}_j) P(u | \mathcal{U}_i) P(e | \mathcal{F}_j) P(r | \mathcal{U}_i, \mathcal{F}_j)]^b}.$$

其中 b 是一个介于0到1之间的模拟退火参数, $P(r | \mathcal{U}_p, \mathcal{F}_q)$ 满足 K 维正态分布, i, p 表示用户隐类编号, j, q 表示服务情境隐类编号;

(3)M步使用E步计算得到的用户隐类和服务情境隐类的联合后验概率重新估算模型参数,包括每一个用户隐类和每一个服务情境隐类的先验概率 $P(\mathcal{U}_p)$ 和 $P(\mathcal{F}_q)$,每一个用户在给定不同用户隐类时出现的条件概率 $P(u | \mathcal{U}_p)$,每一则服务情境在给定不同服务情境隐类时出现的条件概率 $P(e | \mathcal{F}_q)$,同时给定不同的用户情境隐类和服务情境隐类时 K 个指标偏好的正态分布均值 $\mu_{k, \mathcal{U}_p, \mathcal{F}_q}$ 和方差 $\sigma_{k, \mathcal{U}_p, \mathcal{F}_q}^2$:

$$\begin{aligned} P(\mathcal{U}_p) &= \frac{\sum_l \sum_j P(\mathcal{U}_p, \mathcal{F}_j | u_{(l)}, e_{(l)}, r_{(l)})}{L}, \quad P(\mathcal{F}_q) = \frac{\sum_l \sum_i P(\mathcal{U}_i, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}{L}, \\ P(u | \mathcal{U}_p) &= \frac{\sum_{l: u_{(l)}=u} \sum_j P(\mathcal{U}_p, \mathcal{F}_j | u_{(l)}, e_{(l)}, r_{(l)})}{L * P(\mathcal{U}_p)}, \quad P(e | \mathcal{F}_q) = \frac{\sum_{l: e_{(l)}=e} \sum_i P(\mathcal{U}_i, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}{L * P(\mathcal{F}_q)}, \\ \mu_{k, \mathcal{U}_p, \mathcal{F}_q} &= \frac{\sum_l r_{k(l)} P(\mathcal{U}_p, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}{\sum_l P(\mathcal{U}_p, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}, \quad \sigma_{k, \mathcal{U}_p, \mathcal{F}_q}^2 = \frac{\sum_l (r_{k(l)} - \mu_{k, \mathcal{U}_p, \mathcal{F}_q})^2 P(\mathcal{U}_p, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}{\sum_l P(\mathcal{U}_p, \mathcal{F}_q | u_{(l)}, e_{(l)}, r_{(l)})}. \end{aligned}$$

其中 l 表示历史训练样本的编号, L 表示训练样本的总数, i, p 表示用户隐类的标号, j, q 表示服务情境隐类的编号, u, e 表示单个用户和单则服务情境, $\mathcal{U}, \mathcal{F}$ 分别表示用户隐类和服务情境隐类, k 表示QoS指标的编号;

(4)检查模型参数是否收敛,若收敛,结束并保存模型参数,若不收敛,返回(2)执行。

6. 根据权利要求5所述的基于隐语义概率模型的个性化服务推荐方法,其特征在于:步骤6中利用已训练的隐语义概率模型预测用户指标偏好的实现方式如下:

(1)用户 u^t 登陆服务推荐系统,提供需要进行服务推荐的服务情境 e^t ;

(2)推荐系统获取用户的个人数据及其提供的服务情境信息,用下面的公式对该用户在该服务情境下的各个QoS指标偏好进行独立预测:

$$R_k(e^t, u^t) = \int_0^1 r_k \frac{P(e^t, u^t, r_k)}{\int_0^1 P(e^t, u^t, r_k) dr_k} dr_k;$$

其中:

$$P(e^t, u^t, r_k) = \sum_{i,j} P(\mathcal{U}_i) P(\mathcal{F}_j) P(u^t | \mathcal{U}_i) P(e^t | \mathcal{F}_j) P(r_k | \mathcal{U}_i, \mathcal{F}_j)$$

$$P(r_k | \mathcal{U}_i, \mathcal{F}_j) = \frac{1}{\sqrt{2\pi}\sigma_{k,\mathcal{U}_i,\mathcal{F}_j}} \exp \left[-\frac{(r_k - \mu_{k,\mathcal{U}_i,\mathcal{F}_j})^2}{2\sigma_{k,\mathcal{U}_i,\mathcal{F}_j}^2} \right].$$

其中 u^t 、 e^t 分别表示待预测指标偏好的用户及其所处的服务情境, k 表示QoS指标的编号, $R_k(e^t, u^t)$ 表示用户 u^t 在服务情境 e^t 下对第 k 个指标偏好的预测值, $\mathcal{U}$ 、 $\mathcal{F}$ 分别代表用户隐类和服务情境隐类, i 表示用户隐类的标号, j 表示服务情境隐类的编号, $\mu_{k,\mathcal{U}_i,\mathcal{F}_j}$ 和 $\sigma_{k,\mathcal{U}_i,\mathcal{F}_j}^2$ 分别表示给定用户情境隐类 $\mathcal{U}_i$ 和服务情境隐类 $\mathcal{F}_j$ 时第 k 个指标偏好的正态分布均值 $\mu_{k,\mathcal{U}_i,\mathcal{F}_j}$ 和方差 $\sigma_{k,\mathcal{U}_i,\mathcal{F}_j}^2$;最后预测得到用户 u^t 在该服务情境 e^t 下对各个QoS指标的偏好权重为 $(R_1(e^t, u^t), \dots, R_K(e^t, u^t))$,记作 $(\hat{r}_1, \hat{r}_2, \dots, \hat{r}_K)$ 。

7. 根据权利要求6所述的基于隐语义概率模型的个性化服务推荐方法,其特征在于:步骤7中根据预测出的用户个性化的QoS指标偏好,对候选服务进行筛选的过程如下:用户 u^t 在服务情境 e^t 下已得到多个功能相似的候选服务,假设每个服务的各QoS指标上的性能可有其他方法得到,记为 $(q_1, q_2, \dots, q_K)$,每一个服务的总得分采用 $R = \hat{r}_1 \cdot q_1 + \hat{r}_2 \cdot q_2 + \dots + \hat{r}_K \cdot q_K$ 表示,最后根据各个服务的总得分给出服务的排序,作为服务推荐的依据。

一种基于隐语义概率模型的个性化服务推荐系统及方法

技术领域

[0001] 本发明属于服务计算技术领域,具体涉及一种基于隐语义概率模型的个性化服务推荐系统及方法。

背景技术

[0002] 随着互联网技术的飞速发展,服务计算技术得到了广泛应用,Web服务就是这样一种分布运行于Internet之上、支持不同平台之间互操作的松耦合软件系统,它主要通过“发布-查找-绑定”的模式允许服务使用者和提供者之间形成松散的绑定关系,这为服务的使用奠定了基础。但web服务的使用者和提供者相分离的特性,增加了服务使用者理解服务的难度,同时随着Internet上运行的web服务数量不断增多,服务使用者需要从众多功能相似的服务中选出最符合自身需求的一个或一组服务,这对大多数缺乏专业知识的服务使用者来说无疑是一项繁重的任务,所以发展有效的服务推荐技术是服务选择的必然需求。

[0003] 经过对现有技术的检索发现,中国专利申请号200710162463.8,记载了一种自适应服务推荐设备,该装置主要包括:语义分析装置、服务选择装置、服务推荐装置;语义分析装置用于对用户的查询进行语义上的分析;服务选择装置用于找出与语义分析后的查询对应的选择的服务,并根据选择的服务更新服务相关数据库;最后服务推荐装置用于利用获取的选择的服务来查找服务相关数据库以向用户推荐相关服务。

[0004] 进一步检索发现,中国专利申请号200910236492.3,记载了一种个性化服务推荐系统和方法,该方法主要包括:用户信息收集器监控在终端进行的各种操作信息,并进行预处理后存入用户信息数据库,若用户信息数据库方法更新,则启动用户行为分析器进行分析;用户行为分析器扫描用户信息数据库,提取新的用户信息并计入资源信息数据库,计算新的推荐策略并记入推荐策略数据库;上下文感知处理器感知用户的当前上下文,输出当前的上下文描述信息,启动个性化推荐处理器;个性化推荐处理器接收来自上下文感知处理器的消息后,获得当前上下文信息,通过检索推荐策略数据库,获得匹配的左右推荐策略,并按照左右推荐策略通过检索资源信息数据库,匹配合适的资源信息,实时生成个性化推荐服务。

[0005] 进一步检索发现,中国专利申请号20121014234.2记录了一种基于上下文感知和用户偏好的空间信息服务匹配方法,在用户的服务请求和候选空间信息服务的功能性匹配和非功能性匹配的基础上,考虑上下文敏感的用户偏好,计算出用户服务需求与智能空间中的各候选空间信息服务的匹配度,然后依据匹配度将候选空间信息服务推荐给用户;虽然用户可能具有固定的或者重复的偏好,但这些偏好不是在任何时候都相关,该发明基于上下文感知和用户偏好的空间信息服务匹配方法从上下文的角度对用户偏好进行精简,剔除与用户无关的用户偏好;该发明从功能匹配和非功能匹配两个方面进行匹配度计算,提高空间信息服务匹配的准确率。

[0006] 进一步检索发现,中国专利申请号201210253884.2记录了一种用于Web服务推荐的个性化搜索方法,包括以下步骤:步骤1,预处理WSDL文档:通过去除停用词和提取词干两

个预处理步骤,形成词袋;步骤2,抽取用户兴趣:使用改进的TF-IDF公式计算词袋中的每一个词的权重,并乘以该词的时间衰减因子,得到新的权重;选择权重由大至小前k个词作为用户的兴趣词,以及每个词的对应权重,组成k维的用户兴趣向量;步骤3,计算兴趣相似度:设定相似度阈值,超过阈值的用户入选为目标用户的邻居用户;步骤4,排序服务检索结果,根据邻居用户的相似度及其选择服务的次数计算服务的推荐预测值,并将检索结果按照推荐预测值降序排列,从而得到个性化搜索结果。

[0007] 上述的方法都涉及了个性化的服务推荐,但很少针对Web服务的非功能属性筛选服务以满足用户的个性化需求,服务的非功能属性往往是服务性能的主要体现,所以如何利用服务的非功能属性客观地评价服务的性能同时又能满足不同用户的个性化需求是服务推荐需要解决的问题。

发明内容

[0008] 本发明技术解决问题:针对现有技术存在的不足,提供一种基于隐语义概率模型的个性化服务推荐系统及方法,旨在为用户在不熟悉的服务情境下提供指标偏好预测,从而完成基于多维QoS指标的服务综合排序,为用户提供个性化的服务推荐结果。

[0009] 本发明的技术解决方案:一种个性化服务推荐中基于隐语义概率模型的用户指标偏好预测方法,具体步骤如下:

[0010] 步骤1、确定评价服务性能优劣的服务QoS指标体系

[0011] 所述的服务QoS指标体系是指整个Web服务系统统一采用的用于评价一系列功能相似服务性能优劣所用QoS指标的集合,不同的系统可以根据需要选取适当的QoS指标组成自己的QoS指标体系用于评价服务的性能优劣;

[0012] 步骤2、建立用户、用户指标偏好以及服务情境三者之间的隐语义概率模型

[0013] 所述的用户指标偏好是指用户对步骤1中所述服务QoS指标体系中各个QoS指标的偏好程度,对每个QoS指标的偏好程度值介于0到1之间,并且对各个QoS指标的偏好值总和为1;所述的服务情境是指在何种场景下使用何种功能的服务,每一个服务情境 e 用三元组 (w_1, w_2, w_3) 表示,其中 w_1 表示服务完成的与业务无关的基本功能,如视频功能、导航功能等, w_2 表示用户使用该功能的服务完成的具体业务活动,如学术会议、军事导航等, w_3 表示用户调用服务的终端设备,如手机、PC机等;所述的用户、用户指标偏好、服务情境三者之间的隐语义概率模型是指单个用户以不同的概率依赖不同的用户隐类而存在,单则服务情境以不同的概率依赖不同的服务情境隐类存在,指标偏好同时以不同的概率同时依赖不同的用户隐类和服务情境隐类存在的概率模型,模型的图像化表示如7所示。

[0014] 所述的用户隐类是指非人为事先确定的而是从历史经验数据中学习得到的用户聚簇;所述的服务情境隐类是指非人为事先确定的而是从历史经验数据中学习得到的服务情境聚簇;

[0015] 步骤3、收集不同用户在不同的服务情境下使用不同功能服务时自主提供的指标偏好信息,作为历史经验数据存入数据库,为训练步骤2中建立的隐语义概率模型的参数作准备,存储格式为(用户,服务情境、指标偏好)三元组;

[0016] 步骤4、用EM算法及已收集的历史经验数据训练隐语义概率模型的参数

[0017] 所述的隐语义概率模型的参数是指所有用户隐类 $\{U_1, U_2, \dots, U_I\}$ 的先验概率 P

(U_i) ($1 \leq i \leq I$)、服务情境隐类 $\{E_1, E_2, \dots, E_J\}$ 先验概率 $P(E_j)$ ($1 \leq j \leq J$)、给定一个用户隐类 U_i 的情况下单个用户 u 出现的条件概率 $P(u|U_i)$ ($1 \leq i \leq I$)、给定一个服务情境隐类 E_j 的情况下单则服务情境 e 出现的条件概率 $P(e|E_j)$ ($1 \leq j \leq J$)、给定用户隐类 U_i 和服务情境隐类 E_j 的情况下用户指标偏好向量 r 出现的概率 $P(r|U_i, E_j)$ ($1 \leq i \leq I, 1 \leq j \leq J$);其中 I, J 分别表示用户隐类和服务情境隐类的总个数, i, j 分别表示用户隐类和服务情境隐类的编号;

[0018] 步骤5、获得用户需要服务推荐的请求,包括个人信息、服务推荐所依赖的服务情境;

[0019] 步骤6、用已训练的隐语义概率模型预测指定用户在特定服务情境下的未知指标偏好;

[0020] 步骤7、根据预测出的用户个性化的QoS指标偏好,对候选服务进行综合筛选,从而选出最贴近该用户需求的服务,将推荐结果返回给用户。

[0021] 所述步骤3中收集历史经验数据的过程如下:

[0022] (1)用户在其熟悉的服务情境下调用熟悉的web服务时可以直接与系统交互,使用层次分析法给出自己给出对各个指标偏好的两两比较结果,建立比较矩阵;

$$[0023] \quad C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix},$$

[0024] (2)验证比较矩阵的一致性是否在可接受的范围内,若可接受,进入(3),否则返回(1);

[0025] (3)计算比较矩阵的最大特征值对应的特征向量,将该特征向量归一化后所得到的就是用户对各个QoS指标的偏好向量 W_{QoS} ,将(用户、服务情境、指标偏好)三元组存入数据库。

[0026] 所述步骤4中隐语义概率模型的参数训练过程如下:

[0027] (1)从数据库中取出所有的用户、服务情境、指标偏好三元组作为训练数据,是用户在其熟悉的服务情境下利用层次分析法给出的个性化指标偏好权值,每个三元组(u, e, r)表示用户 u 在服务情境 e 下对服务的各个QoS指标的偏好权重为 r , r 是一个 K 维向量($r_1, r_2, \dots, r_K$),为简单起见,假设权重向量 r 的每一维相互独立并且服从正态分布,则联合概率 $P(u, e, r)$ 可用下面两个公式表示,

$$[0028] \quad P(u, e, r) = \sum_i \sum_j P(U_i) P(E_j) P(u|U_i) P(e|E_j) P(r|U_i, E_j),$$

[0029] 其中 $P(r|U_i, E_j)$ 为 K 维正态分布, i, j 分别表示用户隐类和服务情境隐类编号;

[0030] (2)E步计算隐类的联合后验概率,

$$[0031] \quad P(U_p, E_q | u, e, r) = \frac{[P(U_p) P(E_q) P(u|U_p) P(e|E_q) P(r|U_p, E_q)]^b}{\sum_i \sum_j [P(U_i) P(E_j) P(u|U_i) P(e|E_j) P(r|U_i, E_j)]^b},$$

[0032] 其中 b 是一个介于0到1之间的模拟退火参数, $P(r|U_p, E_q)$ 为 K 维正态分布, p, q 表示用户隐类和服务情境隐类编号;

[0033] (3)M步使用E步计算得到的后验概率重新估算模型参数,包括每一个用户隐类和

每一个服务情境隐类的先验概率 $P(U_p)$ 和 $P(E_q)$,每一个用户在给定不同用户隐类时出现的条件概率 $P(u|U_p)$,每一则服务情境在给定不同服务情境隐类时出现的条件概率 $P(e|E_q)$,同时给定不同的用户情境隐类和服务情境隐类时K个指标偏好的正态分布均值 $\mu_{k|U_p,E_q}$ 和方差 $\sigma_{k|U_p,E_q}^2$:

$$[0034] \quad P(U_p) = \frac{\sum_l \sum_j P(U_p, E_j | u_{(l)}, e_{(l)}, r_{(l)})}{L}, P(E_q) = \frac{\sum_l \sum_i P(U_i, E_q | u_{(l)}, e_{(l)}, r_{(l)})}{L},$$

$$[0035] \quad P(u|U_p) = \frac{\sum_{i: u_{(i)}=u} \sum_j P(U_p, E_j | u_{(i)}, e_{(i)}, r_{(i)})}{L * P(U_p)}, P(e|E_q) = \frac{\sum_{i: e_{(i)}=e} \sum_i P(U_i, E_q | u_{(i)}, e_{(i)}, r_{(i)})}{L * P(E_q)},$$

$$[0036] \quad \mu_{k|U_p,E_q} = \frac{\sum_l r_{k(l)} P(U_p, E_q | u_{(l)}, e_{(l)}, r_{(l)})}{\sum_l P(U_p, E_q | u_{(l)}, e_{(l)}, r_{(l)})}, \sigma_{k|U_p,E_q}^2 = \frac{\sum_l (r_{k(l)} - \mu_{k|U_p,E_q})^2 P(U_p, E_q | u_{(l)}, e_{(l)}, r_{(l)})}{\sum_l P(U_p, E_q | u_{(l)}, e_{(l)}, r_{(l)})}.$$

[0037] 其中l表示历史训练样本的编号,i,p表示用户隐类的标号,j,q表示服务情境隐类的编号,u,e表示单个用户和单则服务情境,U,E表示用户隐类和服务情境隐类,k表示QoS指标的编号。

[0038] (4)检查模型参数是否收敛,若收敛,结束并保存模型参数,若不收敛,返回(2)执行。

[0039] 所述步骤6中利用已训练的隐语义概率模型预测用户指标偏好的实现方式如下:

[0040] (1)用户 u^t 登陆服务推荐系统,提供需要进行服务推荐的服务情境 e^t ;

[0041] (2)推荐系统获取用户的个人数据及其提供的服务情境信息,用下面的公式对该用户在该服务情境下的各个QoS指标偏好进行独立预测

$$[0042] \quad R_k(e^t, u^t) = \int_0^1 r_k \frac{P(e^t, u^t, r_k)}{\int_0^1 P(e^t, u^t, r_k) dr_k} dr_k$$

[0043] 其中

$$[0044] \quad P(e^t, u^t, r_k) = \sum_{i,j} P(U_i) P(E_j) P(u^t | U_i) P(e^t | E_j) P(r_k | U_i, E_j)$$

$$[0045] \quad P(r_k | U_i, E_j) = \frac{1}{\sqrt{2\pi}\sigma_{k|U_i,E_j}} \exp \left[-\frac{(r_k - \mu_{k|U_i,E_j})^2}{2\sigma_{k|U_i,E_j}^2} \right].$$

[0046] 最后预测得到用户 u^t 在该服务情境 e^t 下对各个QoS指标的偏好权重为 $(R_1(e^t, u^t), \dots, R_K(e^t, u^t))$,简单记作 $(\hat{r}_1, \hat{r}_2, \dots, \hat{r}_K)$ 。

[0047] 所述步骤7中根据预测出的用户个性化的QoS指标偏好,对候选服务进行筛选的过程如下:用户 u^t 在服务情境 e^t 下已得到多个功能相似的候选服务,假设每个服务的各QoS指标上的性能可有其他方法得到,记为 $(q_1, q_2, \dots, q_K)$,那么每一个服务的总得分采用 $R = \hat{r}_1 \cdot q_1 + \hat{r}_2 \cdot q_2 + \dots + \hat{r}_K \cdot q_K$ 表示,最后根据各个服务的总得分给出服务的排序,作为服务推

荐的依据。

[0048] 本发明与现有技术相比的优点在于：本发明提供的技术方案建立了一种有关用户、用户指标偏好、服务情境之间的概率依赖模型，用历史数据对模型训练以后，可以利用该模型方便地预测出特定用户在陌生服务情境下的指标偏好向量，最后结合服务的非功能属性和已经预测的用户指标偏好对候选服务进行综合评价以实现个性化服务推荐。已有方法一般只简单地将候选服务的各维度QoS值进行加权平均，据此给出候选服务的综合排序，权值一般是先确定的，显然无法体现用户在服务调用过程中的个性化需求。本专利提出的基于隐语义概率模型的个性化服务推荐方法可以以一种自动的方式预测出用户在特定情境下的个性化指标偏好，即使在用户自己也无法明确表达自己的偏好信息的时候。同时本发明的模型训练可以离线进行，所以提高了在线服务推荐的效率。

附图说明

- [0049] 图1是本发明系统的功能框图；
[0050] 图2为图1中历史信息收集模块的实现流程图；
[0051] 图3为图1中隐语义概率模型参数训练模块的实现流程图；
[0052] 图4为图1中服务推荐请求模块的实现流程图；
[0053] 图5为图1中个性化指标偏好预测模块的实现流程图；
[0054] 图6为图1中个性化服务推荐模块的实现流程图；
[0055] 图7为隐语义概率模型的图像化表示。

具体实施方式

[0056] 如图1所示，本发明一种基于隐语义概率模型的个性化服务推荐系统及方法由历史信息收集模块、隐语义概率模型参数训练模块、服务推荐请求模块、个性化指标偏好预测模块、个性化服务推荐模块组成。

[0057] 整个实现过程如下：

[0058] 步骤1、确定评价服务性能优劣的服务QoS指标体系

[0059] 所述的服务QoS指标体系是指整个Web服务系统统一采用的用于评价一系列功能相似服务性能优劣所用QoS指标的集合，不同的系统可以根据需要选取适当的QoS指标组成自己的QoS指标体系用于评价服务的性能优劣；

[0060] 步骤2、建立用户、用户指标偏好以及服务情境三者之间的隐语义概率模型

[0061] 所述的用户指标偏好是指用户对步骤1中所述服务QoS指标体系中各个QoS指标的偏好程度，对每个QoS指标的偏好程度值介于0到1之间，并且对各个QoS指标的偏好值总和为1；所述的服务情境是指在何种场景下使用何种功能的服务，每一个服务情境 e 用三元组 (w_1, w_2, w_3) 表示，其中 w_1 表示服务完成的与业务无关的基本功能，如视频功能、导航功能等， w_2 表示用户使用该功能的服务完成的具体业务活动，如学术会议、军事导航等， w_3 表示用户调用服务的终端设备，如手机、PC机等；所述的用户、用户指标偏好、服务情境三者之间的隐语义概率模型是指单个用户以不同的概率依赖不同的用户隐类而存在，单则服务情境以不同的概率依赖不同的服务情境隐类存在，指标偏好同时以不同的概率同时依赖不同的用户隐类和服务情境隐类存在的概率模型，模型的图像化表示如图7所示。图中U代表用户隐类，

E代表服务情境隐类,u、e分别代表单个用户和单个服务情境,r代表指标偏好向量,图7显示了这几个要素之间的概率依赖关系,用户u、服务情境e、指标偏好向量r对用户隐类U和服务情境隐类E的依赖度的大小以条件概率 $P(u|U)$ 、 $P(e|E)$ 、 $P(r|U,E)$ 的大小来衡量。

[0062] 所述的用户隐类是指非人为事先确定的而是从历史经验数据中学习得到的用户聚簇;所述的服务情境隐类是指非人为事先确定的而是从历史经验数据中学习得到的服务情境聚簇;

[0063] 步骤3、收集不同用户在不同的服务情境下使用不同功能服务时自主提供的指标偏好信息,作为历史经验数据存入数据库,为训练步骤2中建立的隐语义概率模型的参数作准备,存储格式为(用户,服务情境、指标偏好)三元组;

[0064] 步骤4、用EM算法及已收集的历史经验数据训练隐语义概率模型的参数

[0065] 所述的隐语义概率模型的参数是指所有用户隐类 $\{U_1, U_2, \dots, U_I\}$ 的先验概率 $P(U_i) (1 \leq i \leq I)$ 、服务情境隐类 $\{E_1, E_2, \dots, E_J\}$ 先验概率 $P(E_j) (1 \leq j \leq J)$ 、给定一个用户隐类 U_i 的情况下单个用户u出现的条件概率 $P(u|U_i) (1 \leq i \leq I)$ 、给定一个服务情境隐类 E_j 的情况下单则服务情境e出现的条件概率 $P(e|E_j) (1 \leq j \leq J)$ 、给定用户隐类 U_i 和服务情境隐类 E_j 的情况下用户指标偏好向量r出现的概率 $P(r|U_i, E_j) (1 \leq i \leq I, 1 \leq j \leq J)$;i,j分别代表什么,其中I,J分别表示用户隐类和服务情境隐类的总个数,i,j分别表示用户隐类和服务情境隐类的编号;

[0066] 步骤5、获得用户需要服务推荐的请求,包括个人信息、服务推荐所依赖的服务情境;

[0067] 步骤6、用已训练的隐语义概率模型预测指定用户在特定服务情境下的未知指标偏好;

[0068] 步骤7、根据预测出的用户个性化的QoS指标偏好,对候选服务进行综合筛选,从而选出最贴近该用户需求的服务,将推荐结果返回给用户。

[0069] 上述各模块的具体实现过程如下:

[0070] 1.历史信息收集模块

[0071] 实现过程如图2所示:收集用户的个人信息,包括姓名、职业、年收入、兴趣爱好等,为每个用户分配一个编号;收集用户提供指标偏好所处的服务情境信息,服务情境表示为三元组(服务基本功能,目标活动,终端设备),比如(导航功能,日常出行,手机)或者(语音功能,学术会议,个人电脑)等等。假设实际确定的各个QoS维度为:可靠性、响应时间、可用性、吞吐量、价格,那么用户在该服务情境下给出的比较矩阵就为5*5的矩阵C,计算其最大特征值对应的特征向量并归一化,即为指标偏好,将用户,服务情境,指标偏好三元组存入数据库。

[0072] 2、隐语义概率模型参数训练模块

[0073] 实现过程如图3所示:将数据库中的历史训练数据取出,然后人工设置用户隐类和服务隐类的个数,假设都为5个,以及隐语义概率模型的初始参数:5个用户隐类的初始先验概率,可均设为1/5;5个服务情境隐类的概率,可均设为1/5;每一个用户在给定不同用户隐类的条件概率初值;每一则服务情境在给定不同服务情境隐类的条件概率初值;每一个指标偏好在给定不同用户隐类和服务情境隐类时的正态分布均值和方差的初值,然后用EM算法结合训练数据迭代训练模型的参数直至收敛并将参数保存。

[0074] 3.服务推荐请求模块

[0075] 实现过程如图4所示,用户登录服务推荐系统,假设该用户已经是系统的注册用户,并且有特定的登录名标识,记为 u^t ,他提供了需要进行服务推荐的服务情境记为 e^t ,具体含义是(汇率转换,网络购物,个人电脑),该模块会记录下 u^t 和 e^t ,提交给个性化指标偏好预测模块和个性化服务推荐模块。

[0076] 4.个性化指标偏好预测模块,实现过程如图5所示。

[0077] 该模块需要获取已训练好的模型参数和用户提供的服务推荐请求信息,基于上面提出的指标偏好预测方法预测出用户指标偏好,假设预测的指标偏好为 $r=(0.3,0.1,0.2,0.3,0.1)$ 。

[0078] 5.个性化服务推荐模块

[0079] 实现过程如图6所示,假设满足汇率转换的服务共有3个 s_1, s_2, s_3 ,它们在五个QoS维度上的归一化值可以由其他方式获得,假设分别记为: $q_1=(0.5,0.5,0.6,0.7,0.6)$, $q_2=(0.8,0.4,0.7,0.2,0.6)$, $q_3=(0.4,0.8,0.2,0.6,0.4)$,那么三个候选服务的得分分别为:

[0080] $g_1=r \cdot q_1=(0.3,0.1,0.2,0.3,0.1) \cdot (0.5,0.5,0.6,0.7,0.6)=0.59$;

[0081] $g_2=r \cdot q_2=(0.3,0.1,0.2,0.3,0.1) \cdot (0.8,0.4,0.7,0.2,0.6)=0.54$;

[0082] $g_3=r \cdot q_3=(0.3,0.1,0.2,0.3,0.1) \cdot (0.4,0.8,0.2,0.6,0.4)=0.46$;

[0083] 所以可得 $g_1>g_2>g_3$,所以给出了服务排序列表为 s_1, s_2, s_3 ,返回给用户供用户选择。

[0084] 本发明未详细阐述部分属于本领域公知技术。

[0085] 以上所述,仅为本发明部分具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本领域的人员在本发明揭露的技术范围内,可轻易想到的变化或替换,都应涵盖在本发明的保护范围之内。

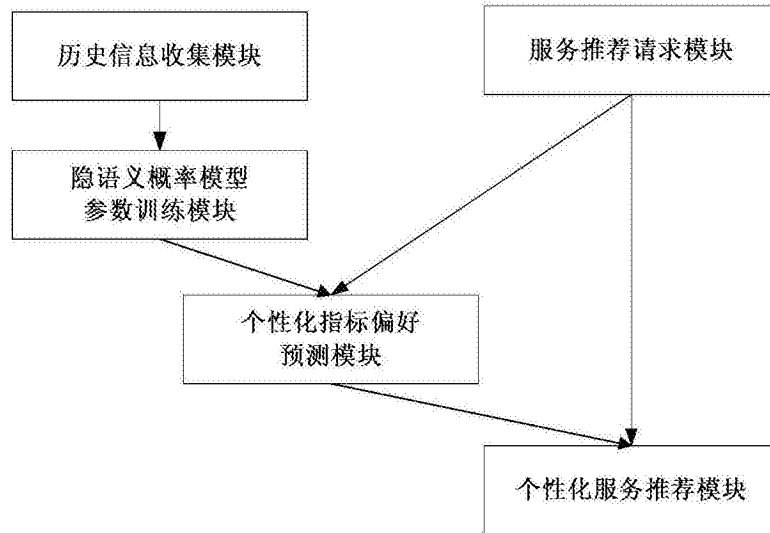


图1

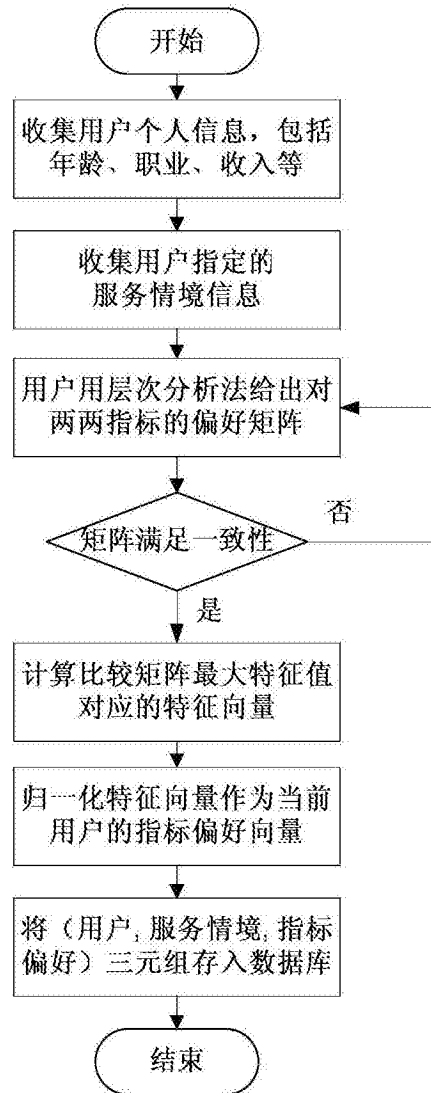


图2

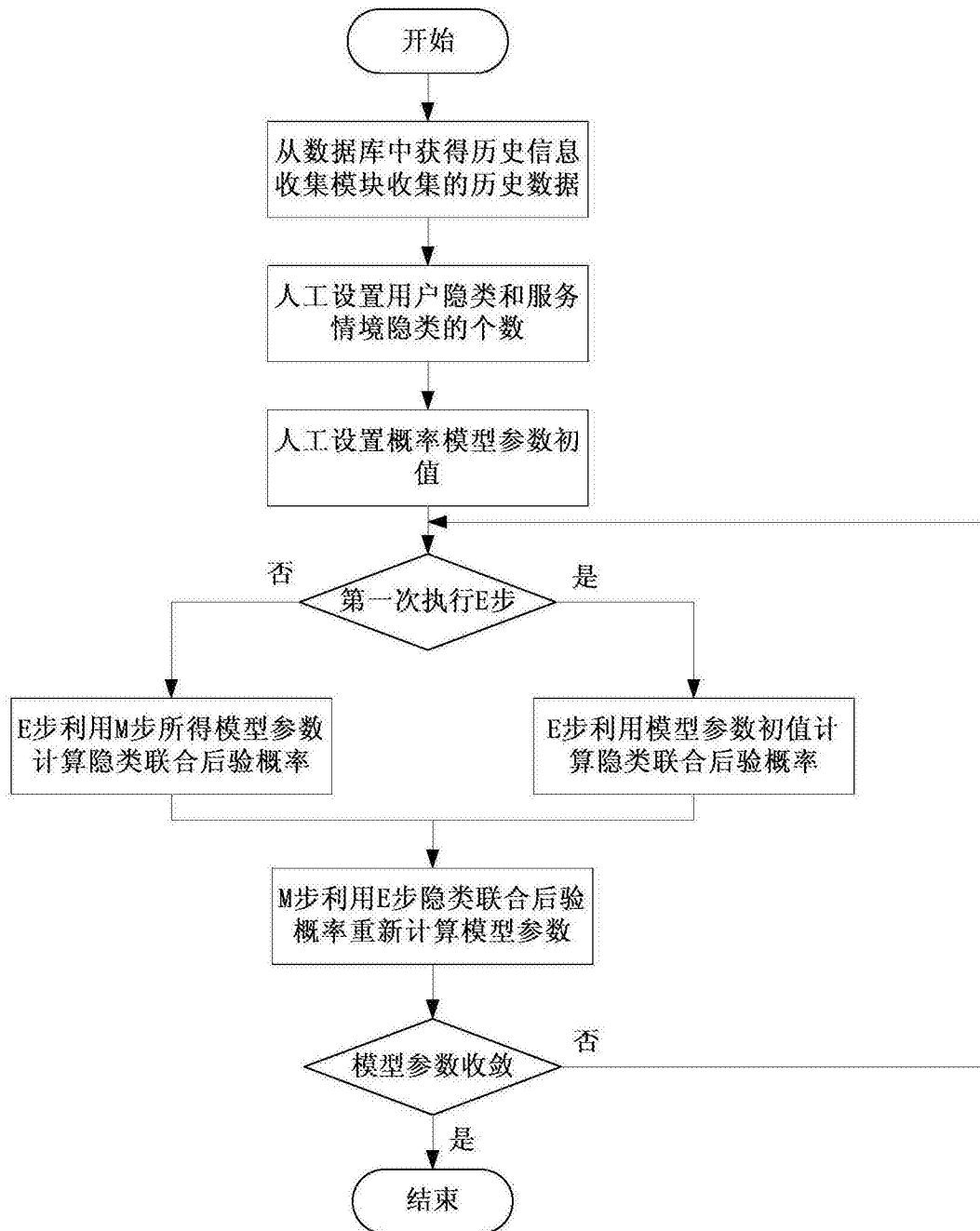


图3

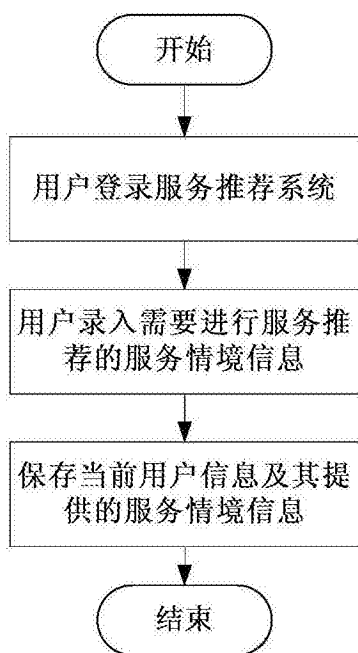


图4

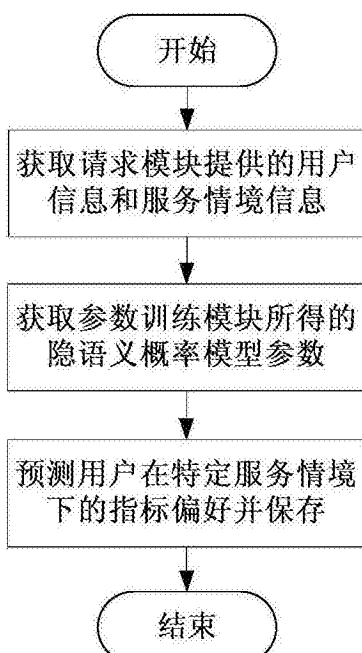


图5

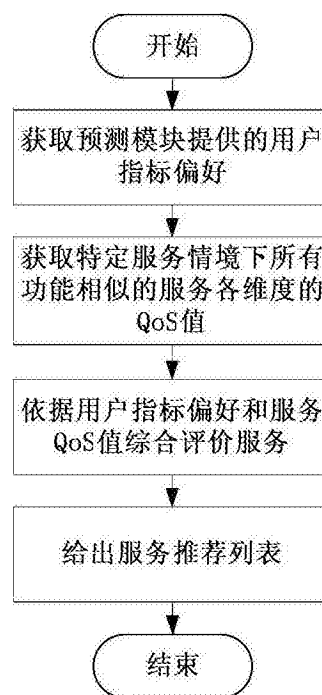


图6

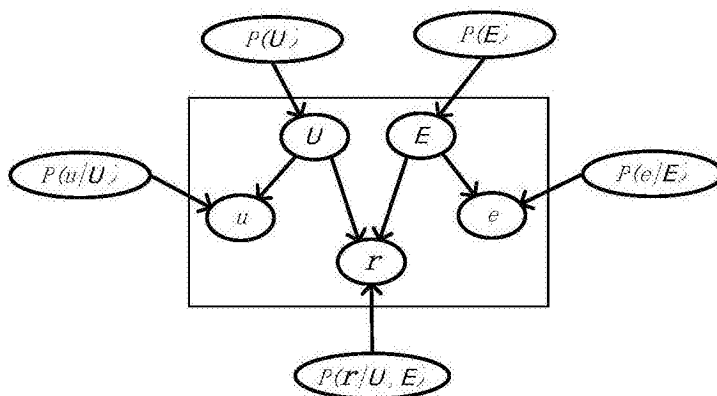


图7