

面向智能搜索引擎的用户兴趣特征模型研究

张英健

(盐城工学院 学报编辑部,江苏 盐城 224051)

摘要:通过直接学习和反馈学习来挖掘智能搜索引擎用户的兴趣特征信息,并建立了兴趣特征库。提出了通过用户兴趣特征增量算法自动修改用户的兴趣特征权值。最后通过实例验证了用户兴趣特征模型在检索中应用的意义。

关键词:智能搜索引擎;用户兴趣特征;特征数据库

中图分类号:TP393.4 **文献标识码:**A **文章编号:**1671-5322(2008)03-0037-04

用户兴趣特征模型是智能搜索引擎实现个性化检索服务的关键。当前,搜索的个性化问题并没有得到解决,不同的人使用相同的检索词得到的检索结果是相同的^[1-2]。其主要原因就是没有对用户在使用搜索引擎时,记录用户的搜索行为,并挖掘用户的兴趣^[3]。因此,识别用户在检索过程中的行为及需求特征,记录、学习并且维持用户的需求偏好,及时了解用户需求信息的更新状况,并实现用户需求管理和主动推送等都给搜索引擎提供了更高的要求,而建立用户兴趣特征模型就可以很好的解决这些问题。

1 用户兴趣特征模型的工作原理

智能搜索引擎系统利用客户端浏览器的用户的兴趣需求选择和用户对检索结果的反馈信息^[4],推理用户在搜索系统内部的兴趣需求取向,系统的流程如图1所示。用户首次使用时,通过注册系统将企业的信息、需求等个性信息输入系统,形成用户特征化数据库,并将这些信息提交系统,以便系统使用。其设计目标是开发一个能够检索网络信息,针对不同用户提供不同检索结果,具有一定智能性的个性化搜索引擎。智能搜索引擎系统应保留每个用户的个人信息,并允许用户对自己的信息进行定制。在使用系统时,系统对用户的行为进行记录,并逐渐形成每个用户的用户模型。在用户检索时,根据信息资源和用户模型,给用户提供更相关度更高的信息资源。

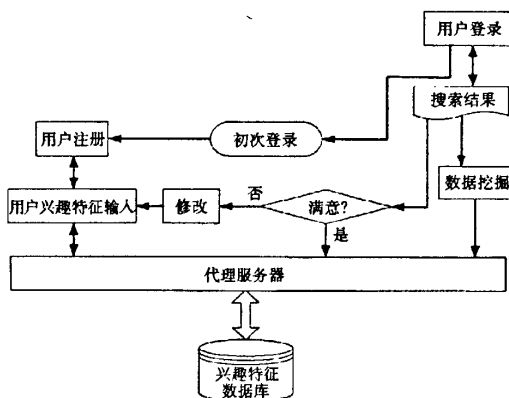


图1 用户兴趣特征模型在搜索引擎中的工作原理

Fig. 1 Working principle of interest feature based search engine

2 用户兴趣特征管理关键技术

2.1 用户兴趣特征挖掘

用户兴趣的学习是指根据用户对浏览信息的选择及使用用户的反馈信息,采取学习方法来逐步明确用户兴趣的一个过程,实质上是一个机器学习的过程,可以采取多种机器学习的方法来实现^[5-6]。本文中用户个性化信息主要采用直接学习和反馈学习的方法来提取。首先根据用户的交互操作来抽取用户的特征兴趣信息,形成用户的兴趣,然后根据搜索的结果,用户进行评价,将评

收稿日期:2008-05-30

作者简介:张英健(1962-),女,江苏泰兴人,副编审,主要研究方向为期刊编辑、计算机应用。

价的结果对生成的用户兴趣模型进行调整。通常一个好的用户兴趣特征的建立需要一定的时间,它需要通过系统和用户之间的不断交互和学习而不断得到完善。由于用户兴趣不是一成不变的,让用户每次访问网站时都输入这些内容会使用户感到很繁琐,而且用户一般不可能提供所有的兴趣以及感兴趣的程度。因此,系统还可以通过用户已访问过的页面去挖掘用户兴趣。通常,用户只会访问自己感兴趣的页面。因此,用户所访问过的页面包含用户感兴趣的信息,通过对这些页面进行分析来更新用户兴趣模型。

2.2 用户兴趣特征管理

用户兴趣特征对用户兴趣进行描述,用户兴趣一般包括了自身信息、关联信息等,根据知识点之间的相近和相关属性,可以对用户兴趣进行扩展。用户可以通过正面的直接建立和修改来管理用户兴趣特征,也可以由用户代理服务器在“黑箱”中暗中操作。

2.3 用户特征兴趣模型生成

在生成用户特征兴趣模型时,用户与系统进行交互,根据交互的内容构成兴趣树,从兴趣树的根节点往下,提供给用户选择的兴趣就越具体。从兴趣树的根节点通往叶节点的过程是发现用户兴趣的,用户个性化信息的提取过程就是从兴趣树中找出一棵用户兴趣子树的过程,如图 2 所示。

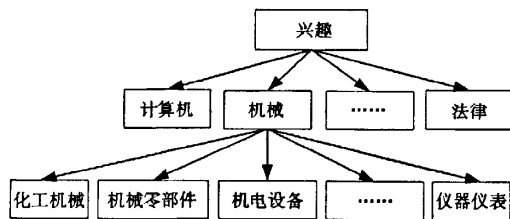


图 2 用户兴趣子树

Fig. 2 User interest sub-tree

2.4 用户个人兴趣特征的权值算法

用户个人兴趣为由一个二元组组成的集合, $U_i = \{(l_{i,1}, \omega_{i,1}), (l_{i,2}, \omega_{i,2}), (l_{i,3}, \omega_{i,3}), \dots, (l_{i,n}, \omega_{i,n})\}$ 其中, $l_{i,j} \in T$, $T = \{t_1, t_2, \dots, t_n\}$ 为词条集, $\omega_{i,j} \in [0, 1]$ 为 $l_{i,j}$ 词条在用户兴趣中的权重, $\sum_{j=1}^n \omega_{i,j} = 1$

用户从兴趣树的根节点向叶子节点方向进行选择,根据用户的选择,生成用户的兴趣模型。用

户的个人特征兴趣生成算法根据用户在兴趣树上的选择结果得到用户的兴趣生成子树,从用户的兴趣生子树可以得到用户个人兴趣特征 U_i 。

步骤 1: 设置兴趣树中叶节点权 $\text{sub}\omega_1 = 1$;

步骤 2: 中间节点权 $\text{mid}\omega_j = \sum_{i=1}^x \omega_i$, 其中 x 为该中间节点的下一级节点数目;

步骤 3: 将所有节点按照权值从大到小排列, 取前 n 个节点, 为 $(l_1, v_1), (l_2, v_2) \dots (l_n, v_n)$, 其中 l_i 为词条, v_i 为对应词条在兴趣树中的权重 v_j

步骤 4: 令 $U_i = \{(l_{i,1}, \omega_{i,1}), (l_{i,2}, \omega_{i,2}), (l_{i,3}, \omega_{i,3}), \dots, (l_{i,n}, \omega_{i,n})\}$, 其中 $(l_{i,2}, \omega_{i,2}) = (l_i, v_i / \sum_{k=1}^n v_k)$, $1 \leq j \leq n$, U_i 即为用户的个人兴趣特征。

2.5 智能搜索引擎用户行为预测算法

在经过多次检索之后,智能搜索引擎用户兴趣特征库自身需要对库内的知识词条的权重进行调整或添加新的词条,智能搜索引擎用户兴趣特征库调整增量算法如下:

步骤 1 计算用户兴趣特征词条权重。

$K(t). \text{Weight} = K(t). \text{Weight} + f_i \times F(n)$ 其中, f_i 表示该用户兴趣特征词条出现的频度, $F(n)$ 表示一单调非递减关联函数,与检索时间紧密有关,即检索的时间越接近当前时间,其值越大。

步骤 2 计算预测信息相关度。

$\text{Value}(l_i) = \sum_Q K(t'_k) \text{Weight} \times R(K(t'_k), K(t''_j))$, Weight 其中 Q 为 $\{(t'_k, t''_j) | (t'_k, f''_k) \in C(l_i), R(K(t'_k), K(t''_j)), \text{Weight}(t'_k, t''_j)) \in \text{Rule}\}$

计算后的 $\text{Value}(l_i)$ 值越大,说明该信息与用户潜在的兴趣的相关性越大。

步骤 3 计算预测推送信息的可信度。

系统预测的用户感兴趣的信息可信度是

$$CT(K) = \frac{\sum_Q K(t'_k) \text{Weight} \times R(K(t'_k), K(t''_j)), \text{Weight}}{\sum_Q K(t'_k) \text{Weight}}$$

当 $CT(K)$ 大于某一设定值时,系统将自动推送该信息至用户界面。

3 用户兴趣代理

代理是指系统中持续自主发挥作用的计算实体(代理服务器),代理具有独立性,自主性和交互性等特性,用户的兴趣特征及用户的个性化搜

索都是通过代理来完成,代理主要负责:

(1)个性化兴趣特征的建立、获取及设置。个性化兴趣特征的建立是用户与代理之间的交互的过程。用户使用个性化的系统时,都是用户用代理进行交互的过程,用户根据用户兴趣树选择兴趣子树,根据用户的个人特征兴趣生成算法生成个性兴趣特征,保存在代理服务器中;同时,用户可以对个性兴趣特征进行修改或重新设置。

(2)对搜索结果过滤。在搜索引擎将检索到的结果返回到代理时,先通过代理中的个性化兴趣特征对信息进行过滤,然后将处理的结果返回给用户的浏览器。

(3)用户在使用搜索引擎时,首先输入关键词,点击搜索时输入的关键词与用户的兴趣特征将同时作为搜索关键词,由搜索引擎将检索出的对应结果返回到用户的浏览器,用户的兴趣特征起到了一个过滤的作用,如图3所示。

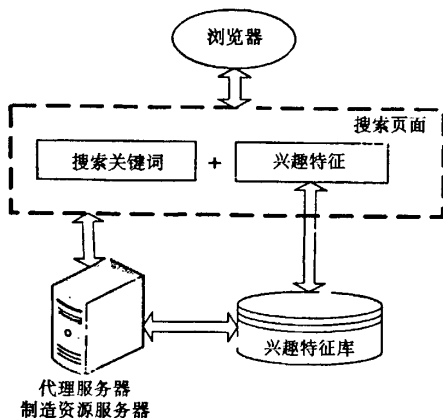


图3 用户兴趣特征的过滤模型

Fig.3 Filter model of user interest feature

4 实例

本文采用“张三”为用户创建用户的兴趣特征,首先进入 internet explorer,打开网络信息搜索页面,点进入个性化搜索页面,因张三为新用户,因此需要新用户注册,点击“新用户注册”链接到

“用户信息注册”页面,输入用户的个人信息后,以及选择用户的研究领域“机械”后提交,将进入用户的子兴趣选择,选择“化工机械”子兴趣,提交后系统生成用户的兴趣特征:主兴趣“机械”,次兴趣“化工机械”;系统根据内部算法自动生成兴趣特征的各权值:“机械”权值1,“化工机械”权值2。若用户对搜索的信息不满意,或者需要修改用户的密码以及更改用户的兴趣,可以通过“密码修改”页面和“兴趣特征”页面修改。此时用户就可以输入想要搜索的关键词进行网络信息搜索。

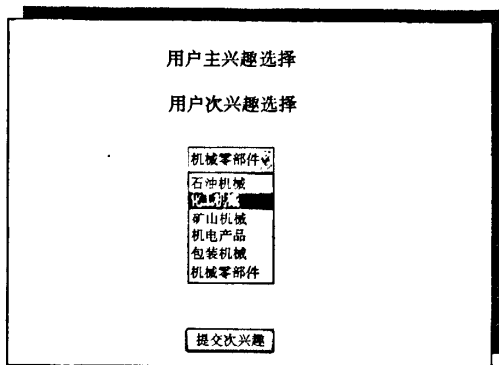


图4 用户兴趣特征选择页面

Fig.4 User interest feature choosing page

用户在使用兴趣特征和未使用兴趣特征所搜索出来的结果是不相同的,而使用了用户兴趣特征的检索结果更精确些。

5 结束语

本文论述了用户兴趣特征在信息搜索中的作用;详细论述了用户兴趣特征的获取方式;论述了用户兴趣特征中权值的算法以及建立用户兴趣特征模型的过程,用户可以通过信息反馈修改用户的兴趣特征;最后通过实例来介绍用户兴趣特征模型在检索中的应用。

参考文献:

- [1] 曾春,刑春晓. 个性化服务技术综述[J]. 软件学报,2002,13(10):1952-1961.
- [2] 王志军,于超. 基于隐式反馈的个人信息检索技术及实现[J]. 计算机工程,2003,29(6):158-159.
- [3] 王斌,李少波,谢庆生. 基于ASP的网络化制造资源智能检索知识库模型设计[J]. 贵州工业大学学报(自然科学)

版),2006,35(2):18-22.

- [4] 李勇,桑艳艳. Web 智能检索中动态相关反馈技术研究[J]. 情报理论与实践,2003,26(2):178-181.
- [5] 王斌,谢庆生. Web 教学资源主题检索系统设计与实现[J]. 现代图书情报技术,2006(1):62-65.
- [6] David A Grossman, Ophir Frieder. Information Retrieval: Algorithms and Heuristics[M]. Kluwer Academic Publishers, 1998.

Study on Intelligent Search Engine Oriented Model of User Interest Features

ZHANG Ying-jian

(Editorial Department, Journal of Yancheng Institute of Technology, Jiangsu Yancheng 224051, China)

Abstract: Intelligent search engine oriented user interest features were mined through direct and feedback study, and an interest feature-related database was established. User interest feature ascending algorithm was proposed to revise the rights of user interest features automatically. Finally, a case was employed to prove the importance of establishing user interest model.

Keywords: intelligent search engine, user interest feature, feature database

(上接第 36 页)

动画,Java Script 特效等技术,使得系统有很好的用户界面,给人一个舒适的操作环境。如果数据库使用 Oracle,相信安全性设计过程会更简单,安全性将更高。

参考文献:

- [1] 邓子云,张赐. JSP 网络编程从基础到实践[M]. 北京:电子工业出版社,2006.
- [2] 余金山,林惠. SQL Server 2000/2005 数据库开发实例入门与提高[M]. 北京:电子工业出版社,2006.
- [3] 彭海云,朱秀丽. 基于 JSP 的高校师生互动平台的构建[J]. 计算机系统应用,2006(10):55-57.
- [4] 柳永坡,刘雪梅,赵长海. JSP 应用开发技术[J]. 北京:人民邮电出版社,2005.
- [5] 王琴,杨宗凯,吴砥. 基于工作流和 JSP/Servlet 技术的网上项目申报与管理系统设计[J]. 计算机应用研究,2006(12):181-184.
- [6] 谢长勇,刘知贵,郑晓红,等. 基于模式 2 的 JSP 管理信息系统的设计与实现[J]. 计算机应用研究,2006(7):135-137.
- [7] 李爱国,邹黎敏,伍俊良. 基于 PHP 信息发布系统实现的新方法[J]. 盐城工学院学报:自然科学版,2008(1):57-61.

Security Design and Implementation of Commodity Marketing System Based on JSP

YANG Jun

(Department of Admission and Employment of Yancheng Institute of Technology, Jiangsu Yancheng 224003, China)

Abstract: This paper aims to describe the feature and superiority of JSP and accomplish commodity marketing system by JSP. The demand analyses and function for the system are given. Especially, we introduce character filter technology, md5 one-way password algorithm, stochastic algorithm, and JSP's session object to achieve security design.

Keywords: JSP; security design; commodity marketing system