第 43 卷 第 9 期 计 算 机 工 程 2017 年 9 月
Vol. 43 No. 9 Computer Engineering September 2017

·人工智能及识别技术· 文章编号：1000-3428(2017)09-0214-06 文献标志码：A 中图分类号：TP311

基于兴趣衰减的个性化排序算法

王 林，刘继源，马安进

（西安理工大学 自动化与信息工程学院，西安 710048）

摘 要： 目前多数个性化排序算法未考虑用户兴趣随时间产生的漂移变化，从而影响排序质量。为此，提出一种融合用户兴趣衰减的个性化排序算法。利用传统个性化排序算法的用户兴趣模型，及用户搜索兴趣的变化规律，分析搜索兴趣程度的时间衰减性，以人类遗忘曲线为基础给出适合搜索兴趣变化的指数遗忘函数，并将其运用到传统个性化排序算法中。实验结果表明，与基于兴趣模型的个性化排序算法相比，该算法能提高个性化搜索引擎的查准率。

关键词： 搜索引擎；个性化排序；搜索兴趣；兴趣漂移；遗忘曲线

中文引用格式： 王 林，刘继源，马安进. 基于兴趣衰减的个性化排序算法[J]. 计算机工程，2017，43(9)：214-219，227.

英文引用格式： WANG Lin，LIU Jiyuan，MA Anjin. Personalization Sorting Algorithm Based on Interest Attenuation[J]. Computer Engineering，2017，43(9)：214-219，227.

Personalization Sorting Algorithm Based on Interest Attenuation

WANG Lin，LIU Jiyuan，MA Anjin

（School of Automation and Information Engineering，Xi'an University of Technology，Xi'an 710048，China）

【Abstract】 The most currentranking algorithm don't take users' interest drifting over time into consideration，which affects the sorting quality. In order to solve this problem，a method ranking algorithm merging user interest attenuation is proposed. In this method，users' interest model of traditional personalization ranking algorithm and changing law of users' searching interests are used to analyze the time attenuation of search interest. The exponential forgetting function fitting searching interest based on human forgetting curve is proposed and applied to the personalized ranking algorithm. Experimental results show that the new algorithm can improve the precision of personalized search engine compared with the ranking algorithm based on users' interest model.

【Key words】 search engine；personalization sorting；search interest；interest drifting；forgetting curve

DOI： 10.3969/j.issn.1000-3428.2017.09.038

0 概述

随着互联网的迅速发展，搜索引擎已成为人们获取信息的重要通道。但由于网上信息资源的爆炸式增长，导致搜索引擎执行一次搜索返回的匹配结果成千上万，而用户真正关心的只是其中的一小部分，只有将这部分网页排在前面才能带给用户有效的搜索体验，这就是排序算法的作用。传统的排序算法仅考虑关键词与网页、网页与网页之间的相关度，返回的排序结果是“面向检索”而非“面向用户”的，如目前常用的 2 类排序算法，基于内容匹配的排序算法[1]和基于链接分析的排序算法[2]。基于内容匹配的排序算法是通过计算查询词对页面主题内容的表征程度进行排序的，以词频位置加权排序算法的应用最为广泛。由于这种排序算法过于依赖词的重要性，网页质量无法得到保证，进而出现了基于链接分析的排序算法，最具代表性的是文献[3]提出的 HITS（Hyperlink-Induced Topic Search）算法、文献[4]提出的 PageRank 算法及文献[5]提出的 Hilltop 算法。但是这些传统排序算法都没有将用户的搜索意图考虑进去，排序结果与用户兴趣的相关性很难得到保证。

为了让排序结果更准确，更接近用户搜索意图，学者们引进了“个性化”的思想，研究面向“用户”的个性化排序算法[6-7]。文献[8]在不同粒度上多次使用 SVD 技术和 K-means 聚类技术对用户浏览历史进行文档聚类和词聚类，建立用户兴趣模型反馈原始排序；文献[9]提取用户的兴趣关键词建立模型，

作者简介： 王 林（1963—），男，教授、博士，主研方向为社交网络、数据挖掘；刘继源，硕士研究生；马安进，硕士。

收稿日期： 2016-08-03 **修回日期：** 2016-10-03 **E-mail：** wanglin@xaut.edu.cn

并根据该模型利用相似度算法计算网页得分进行排序;文献[10]建立网页领域知识库来改进用户兴趣模型,并用余弦相似度算法计算结果网页与兴趣网页之间的余弦距离进行排序。

以上个性化排序算法给出了分析用户浏览历史创建用户兴趣模型来优化排序的思想,相比于传统排序算法在排序结果方面都有一定的优化。

但这些个性化排序算法也普遍存在一个问题:没有考虑用户搜索兴趣在更新周期内随着时间变化的情况,即兴趣漂移。兴趣漂移是指随着时间的推移,用户可能会产生新的兴趣,也可能对过去兴趣不再热衷;即使没有产生新的兴趣,在不同时期用户对同一兴趣的热衷度也不同。用户兴趣判断的准确性直接影响个性化排序的准确性,若不考虑兴趣漂移,排序算法中的用户兴趣模型就不能适应用户兴趣的变化,从而导致排序质量不理想。针对个性化排序算法的上述不足,本文在研究搜索兴趣衰减变化和人类记忆遗忘特点的基础上,提出一种基于兴趣衰减的个性化排序算法。

1　个性化排序算法中的用户兴趣

1.1　个性化排序算法的分析

个性化排序算法是在传统排序算法上,结合用户兴趣特征信息,来提高搜索引擎查准率的一种改进算法。用户通过检索词来搜索相关信息,这一主观性行为在一定程度上也反映了用户想要关注的领域,即用户兴趣;同样,也可以根据用户兴趣来过滤无关的搜索结果,或提高相关领域结果的排名,即通过挖掘用户搜索兴趣,预测用户搜索意图,提高搜索引擎的查准率。因此,用户搜索兴趣判断的准确性严重影响着个性化排序算法的质量。

目前挖掘用户搜索意图比较常用的方法是建立用户兴趣模型,即通过分析用户浏览历史、收藏夹或用户主动提供的一些信息,判断用户经常关注的领域及程度,并以一定的形式量化存储,表征用户兴趣。一般情况下,表示用户兴趣信息的主要特征为网页类型和网页关键词,通过挖掘用户浏览历史挖掘用户经常关注的网页类型及关键词,构成可表达用户潜在搜索意图的用户兴趣模型。用户兴趣模型是个性化排序的核心,记录着用户的个性化特征信息,用户兴趣信息通常采用向量空间模型来表示。

本文将用户兴趣模型表示为:

$$\boldsymbol{C}=((c_1,w_1),(c_2,w_2),\cdots,(c_n,w_n)) \quad (1)$$

向量 $\boldsymbol{C}$ 存在所有兴趣类,这些兴趣类互不重叠。而一个用户的兴趣模型则是该向量的一个子集,包含该用户关注的所有兴趣信息,每个兴趣用兴趣方向和兴趣权值(兴趣程度)来表示,其中,c_i 为第 i 个兴趣方向;w_i 为 c_i 在兴趣模型 w 中的兴趣权值。兴趣权值用来表示用户对某个兴趣方向是否感兴趣及感兴趣的程度,用户兴趣模型的表示方法如图 1 所示。

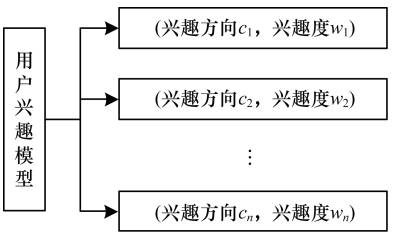

图 1　用户兴趣表示方法

用户使用检索词搜索信息时,搜索引擎会匹配到大量与该检索词有关联的网页,包含各个领域的信息,个性化排序算法可以结合用户兴趣模型,将这些网页根据一定的反馈算法修正网页的得分进行重排序,那么很可能将用户最想获得的信息排在靠前的位置。目前不同个性化排序算法中的反馈算法不同,并没有一个统一的标准,但由于软件的模块化、松耦合特性,只需要找出用户兴趣模型中兴趣权值与网页得分的关系。通过对目前的个性化排序算法进行分析,发现兴趣权值越高,对该兴趣网页的得分正反馈越大。

1.2　搜索兴趣的变化分析

用户每天关注的领域不一定是相同的,则搜索兴趣也并非保持不变。对于用户突然或临时的兴趣,无法预料;但通过对长期或阶段性的兴趣进行分析,那么可获得一些变化规律[11]:

1)用户搜索某类知识时会重复搜索,直到获取一定的知识量后才会减少或停止搜索。这段时间内用户对该知识的搜索频次非常高,间隔时间短,兴趣保持升高的趋势,直到某个程度后兴趣会暂时稳定下来。

2)若用户后期还需要继续搜索该类知识,则会间续搜索。这段时间内的搜索频次相对降低,间隔时间加长,搜索兴趣相对降低。

3)若用户后期可能很少甚至不再进行相关搜索,则说明用户对该类知识的搜索兴趣在下降中,甚至遗忘。

由以上分析可知,用户的搜索兴趣是变化的,而且是有一定规律的。用户兴趣模型是根据用户的历史信息预测用户的搜索意图,即根据历史评

测的兴趣权值来判断用户的兴趣,但由于近因效应存在,即人们识记一系列事务时对末尾部分项目的记忆效果要优于中间部分的项目,而且信息前后间隔时间越长,近因效应越明显,因此历史评测信息距离当前时间越近,则其参考价值就越大。因此,当根据历史评测的兴趣权值判断用户当前的兴趣程度时,应该有一定的衰减,而且衰减程度跟间隔时间是有关系的。

而目前的个性化排序算法没有考虑用户搜索兴趣的衰减情况,这会影响搜索引擎的检索结果:

1)若用户长时间未使用某搜索兴趣,说明用户对该兴趣的关注度在下降,而不考虑衰减就不能准确反映用户对该兴趣关注度的变化情况,会使检索结果偏向于该搜索兴趣。

2)若用户对某搜索兴趣已经遗忘,不考虑衰减就无法准确判断真实的兴趣程度,进而无法从算法中删除。

3)同时还会造成兴趣堆积,用户已经遗忘的搜索兴趣仍会存储在该算法空间中。

因此,不考虑兴趣衰减不仅会造成存储空间的浪费,使模型无限制增大,还会严重影响用户兴趣判断的准确性,从而影响个性化排序的质量。为了改善这一问题,本文分析搜索兴趣的衰减规律,刻画衰减曲线,制定适合搜索兴趣变化的衰减算法,并嵌入到个性化排序算法中,提高搜索引擎的检索质量。

2 搜索兴趣遗忘函数的设计

搜索兴趣,即人感兴趣的事物,属于人的主观意识或记忆。心理学认为兴趣是人们探究某种事物或从事某种活动的心理倾向。在个性化搜索引擎中,不同兴趣的兴趣程度大小与网页的最后得分是相关的,因此,准确判断和跟踪用户的兴趣对个性化排序算法非常有意义。由 1.2 节分析可知,兴趣不仅会增加,也会衰减或遗忘,为了更准确跟踪用户兴趣,考虑兴趣衰减是非常必要的。人的兴趣遗忘是一种自然遗忘的过程,目前已有研究者模拟人类记忆遗忘曲线,跟踪和学习用户兴趣,将其应用到个性化推荐、协同过滤等方面[11-13],并取得很大的进展。因此,本文也借助人类记忆的遗忘规律来研究搜索兴趣的衰减变化,以期提高搜索引擎的查准率。

2.1 兴趣衰减与人类遗忘曲线

德国心理学家 H. Ebbinghaus 经过研究揭示了人类记忆的遗忘规律[13],指出人类记忆时效随时间变化的一般特征,如表 1 所示。并根据他的实验,描绘了遗忘过程的变化曲线,即著名的艾宾浩斯记忆遗忘曲线[14],如图 2 所示。

表 1 人类遗忘规律 %

时间间隔	记忆量
刚记完	100
20 min 后	58
1 h 后	44
8 h~9 h 后	36
1 d 后	33
2 d 后	28
6 d 后	25

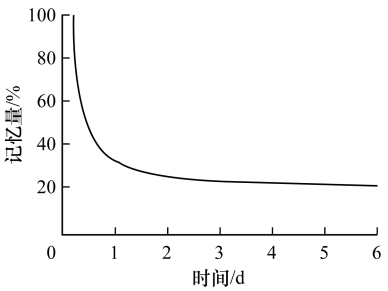

图 2 艾宾浩斯记忆遗忘曲线

从遗忘曲线可以看出,人记忆的遗忘过程呈现先快后慢的变化规律。在记忆过后的初始阶段遗忘速度是最快的,而后逐步减慢,最后以非常缓慢的速度衰减。艾宾浩斯遗忘规律的应用领域很广,文献[15]借鉴艾宾浩斯遗忘曲线来跟踪和学习用户兴趣,通过数学分析工具发现了与遗忘曲线拟合度较高的幂函数曲线,并设计了跟踪用户兴趣变化的基于遗忘曲线的协同过滤推荐算法,提高了个性化推荐的准确性。该幂函数如式(2)所示。

$$Y=31.8\times X^{-0.125},X>0 \tag{2}$$

文献[16]模拟艾宾浩斯遗忘曲线在协同过滤的相似度算法中加入了时间因子,对用户的原始评分进行衰减,以此来反应用户的兴趣变化,改进了传统协同过滤算法,遗忘函数如式(3)所示。

$$f(t,i)=\frac{e^{b}}{(t+t_0)^{c}},c>0,b>0,t_0>0 \tag{3}$$

以上这些改进算法都是模拟艾宾浩斯遗忘曲线跟踪用户兴趣变化,并结合自身算法的一些特点而提出的,即不能直接应用到搜索兴趣的遗忘机制中。搜索兴趣也是人在搜索方面的兴趣,为了更好地跟踪用户搜索兴趣的变化,提高排序质量,本文也模拟人类遗忘规律来表示用户搜索兴趣的遗忘过程,但搜索兴趣也有自己的特点,需要结合人类记忆遗忘曲线的变化规律与搜索兴趣变化自身特点,制定一个合适的遗忘函数 $f(t)$,对历史的用户兴趣信息采取遗忘策略,模拟用户真实的兴趣变化趋势。

2.2　搜索兴趣的指数遗忘函数

模拟用户兴趣遗忘过程的函数主要有指数函数和线性递减函数。由于线性递减函数表示人类兴趣在任何时间段的衰减速率是保持不变的,这与人类自然遗忘规律相悖,即本文不考虑线性递减函数。同时,还要考虑搜索兴趣自身的特点,一般来说,用户搜索并不是时时刻刻的,其周期比较长,需以天为单位;记忆的遗忘是从记忆后就立刻开始进行的,而搜索是人的一种行为,可认为在一定缓冲时间后进行遗忘。因此,根据人类遗忘曲线和搜索兴趣自身特点,提出了一种指数遗忘函数,根据用户访问间隔时间的长短对搜索兴趣权值进行不同程度的衰减,从而把用户兴趣变化考虑到个性化排序算法中。

为方便后文表述,定义相关符号如下:

1)兴趣参考时刻 T_s,指该兴趣的遗忘起始时间。

2)兴趣访问时刻 T,指访问该兴趣的时间。

3)最小遗忘时间间隔 $T_{\min}$,指兴趣开始遗忘时间与兴趣参考时间之差,即遗忘缓冲期。

4)最大遗忘时间间隔 $T_{\max}$,指兴趣度衰减到原始的$\frac{1}{e}$所花费的时间,即衰减周期。

则本文提出的遗忘函数如下:

$$f(t)=\exp\left(-k\frac{t-T_{\min}}{T_{\max}-T_{\min}}\right),T_{\min}\leqslant t\leqslant T_{\max}\quad(4)$$

其中,t 为访问间隔时间,即兴趣访问时间与兴趣参考时间之差(以天为单位),$t=T-T_s$;k 为衰减速率,k 值越大衰减速率越大,本文中定义 $k=1$,可根据实际情况调整。

由式(3)可知,该函数的值域为$[1,\frac{1}{e}]$,单调递减,正好符合人类记忆遗忘曲线收敛性的特点。当访问时间间隔 t 逐渐增大时,函数的值会随之减小,表示兴趣衰减度随之增大。从图 3 可以看出,本文的遗忘函数曲线结合搜索兴趣自身的特点,改变了人类遗忘曲线衰减过快并且记忆后立刻开始遗忘的特性。

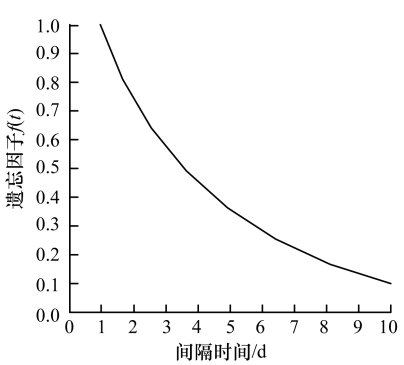

图 3　指数遗忘函数曲线

但式(3)只考虑了搜索兴趣在一个衰减周期($T_{\max}$)内的变化情况,同时还要考虑 $t<T_{\min}$ 和 $t>T_{\max}$ 的情况。本文定义:

1)当 $t<T_{\min}$,即还没有达到最小衰减时间间隔时,定义 $f(t)=1$,表示兴趣度无衰减。

2)当 $t>T_{\max}$,即访问间隔时间超过了一个衰减周期,则可以认为用户暂时失去了这一兴趣,将该兴趣从用户兴趣模型中去除,以免影响用户兴趣判断的准确性。一般情况下,衰减周期时间设置的较长。若在整个衰减周期内都没有关注该兴趣信息,说明用户在这一阶段内对该类网页的兴趣正在下降或失去,用户也可能在将来的某个时间还会关注这一兴趣,但这个时间是未知的,可能会很长,也可能很短。为了尽量减小一个衰减周期都没用关注的兴趣对当前阶段内兴趣模型的影响,将该兴趣从用户兴趣模型中去除是一个比较好的选择。若用户在将来又关注了该兴趣,用户兴趣模型则会重新学习该兴趣,开始一个新的衰减周期。

3　改进的个性化排序算法

传统的个性化排序算法没有考虑兴趣漂移现象,会影响用户兴趣判断的准确性,导致网页排序结果不理想。本文在传统个性化排序算法中加入前面所述的用户兴趣指数遗忘函数,可以得到一种改进的个性化排序算法。在该算法中,用户兴趣在访问时刻的权值是以该兴趣遗忘起始时刻的权值为参考,由指数遗忘函数对其进行修正,降低过去兴趣的权值,实现兴趣度按间隔时间长短进行衰减。

用户每次搜索的主题都是随机的,相邻搜索的主题可能相关,也可能不相关。长期或阶段性的兴趣是指在一段时间内能平稳关注的兴趣,并不是指每次搜索都需要关注的兴趣,这是不现实的。用户会随时出现临时关注的搜索,加入遗忘算法后,由于其权值较低,会很快遗忘掉,降低这种临时兴趣对用户兴趣判断的影响。

设某用户某个兴趣“计算机软件”的兴趣参考时刻权值为 w_1,最小遗忘时间间隔为 3 d,最大遗忘时间间隔为 30 d,则采用传统个性化排序算法和本文改进算法在间隔一段时间后的兴趣权值如表 2 所示。

表 2　权值修正结果

参考权值	访问间隔/d	传统算法权值	本文算法权值
w_1	2	w_1	w_1
w_1	20	w_1	$0.53w_1$
w_1	40	w_1	0

表 2 中,采用传统个性化排序算法的用户兴趣在间隔一段时间后的权值与参考时刻权值一致,仍

为 w_1。而采用本文算法的用户兴趣在间隔一段时间后的权值相对于参考时刻权值有一定程度的衰减，根据遗忘函数公式得：

1）访问间隔 2 d 的修正后权值

$w_1 \cdot f(2) = w_1 \cdot 1 = w_1$

2）访问间隔 20 d 的修正后权值

$$w_1 \cdot f(20) = w_1 \cdot \exp(-\frac{20-3}{30-3}) = 0.53w_1$$

3）访问间隔 40 d 的修正后权值

0

采用遗忘函数后，兴趣权值随时间间隔衰减，模拟用户搜索兴趣的真实变化。基于兴趣衰减的个性化排序算法步骤具体如下：

输入 初步搜索结果集 R（在通用搜索引擎上输入用户查询词产生，并对 R 中前 N 个网页排序），迭代次数初始值 $i=0$

输出 重新排序后的结果集 R_1

1）从用户兴趣模型中获取当前用户的兴趣信息，表示为式（1）。

2）应用式（4），根据兴趣遗忘算法对用户信息进行修正。

3）取第 i 个网页，计算该网页的兴趣类型加成因子：

$$\alpha(t) = \frac{w_t}{\sum_i w_{ti}} \tag{5}$$

其中，分子 w_t 表示网页类型在用户模型中的权值；分母表示用户兴趣模型中所有兴趣类型权值的总和。

4）计算该网页兴趣关键词加成因子 $\alpha(c)$，为网页内容与网页类型下的兴趣关键词序列之间的相似性，首先利用 TF-IDF 计算网页特征词的权重，然后将网页文档与兴趣关键词以向量空间模型表示，则网页文档 d 可表示为：

$$d = \{(k_1, w_{k1}), (k_2, w_{k2}), \cdots, (k_n, w_{kn})\} \tag{6}$$

文档 d 与用户兴趣关键词序列 C 之间的相似度采用余弦相似度计算方法：

$$Sim(d,C) = \frac{d \cdot C}{|d| \cdot |C|} = \frac{\sum_{i=1}^{n} w_{ki} w_{ci}}{\sqrt{\sum_{i=1}^{n} w_{ki}^2} \cdot \sqrt{\sum_{i=1}^{n} w_{ci}^2}} \tag{7}$$

5）计算兴趣加成因子，本文设定为兴趣类型加成因子与兴趣关键词加成因子的和，表示为：

$$\alpha = \alpha(t) + \alpha(c) = \alpha(t) + Sim(d,C) \tag{8}$$

6）计算兴趣得分，为网页原始得分与兴趣加成因子乘积，公式为：

$$Score_i(d) = Score(d) \times \alpha \tag{9}$$

7）如果 $i=N$，转步骤 8）；如果 $i<N$，$i+1$，转步骤 3）。

8）根据兴趣得分大小，降序排列网页并输出 R_1。

算法 1 给出了个性化排序算法的伪代码。

算法 1 基于兴趣衰减的个性化排序算法

输入 R：搜索结果集；C：初始用户兴趣模型；$Score_i(d)$：网页原始得分

输出 R_1：降序排列网页

```
1:w←w · f(t);                    //修正兴趣权值
2:for i←0 to N do                //对每个网页,执行循环
    2.1:compute α(t) of R1 based on Eq.(5);
    //计算兴趣类型加成因子
    2.2:compute α(c) of R1 based on Eq.(7);
    //计算兴趣关键词加成因子
    2.3: compute α of R1 based on Eq.(8);
    //计算兴趣加成因子
    2.4:compute Scorei(d) of R1 based on Eq.(9);
    //计算兴趣得分
    2.5:for i←0 to N-1 do
      2.5.1: for j←0 to N-1-i do
      2.5.2: if Score[j] < Score[j+1] swap
          Score[j] and Score[j+1];
          //降序排列网页
      2.5.3:end for
    2.6:end for
3:end for
```

4 实验结果及分析

本文采用信息检索领域广泛使用的查准率来评价实验结果，定义为检索出的相关文档数在检索出的文档总数中所占的比率。比率越高，搜索引擎的检索精度越高，则排序质量越好。本文重点关注前 10 个结果，定义：

$$\text{查准率} = \frac{\text{前 10 个结果中满足搜索意图的结果数}}{10}$$

为了比较本文算法与传统个性化排序算法在用户兴趣变化时搜索的准确性，采用以下 2 种方法进行重复对比实验并对结果进行统计分析。实验数据选取 CNKI 期刊数据库的 1 000 篇文献摘要，由于期刊数据库的分类比较清晰，能获得较为准确的实验结果，这些摘要在 CNKI 的 10 个不同的二级类中进行随机选择。2 种方法在同一用户下进行 15 次搜索，即起始兴趣相同，搜索内容相同，兴趣变化相同。

方法 1 采用文献[10]提出的个性化排序算法，该算法中没有考虑兴趣的遗忘，兴趣特征为网页类型和关键词。实验中隔一段时间搜索一次，客户端会记录用户每次搜索的行为信息，间隔时间不固定，最后记录搜索结果。

方法 2　在方法 1 采用的算法基础上加入本文提出的兴趣遗忘算法,并进行实验,实验过程与方法 1一致。

然后让用户用 2 种方法以不同领域的关键词进行重复实验,实验记录了 15 个不同领域内搜索的总结果数、每个结果的名次、满足搜索意图的总结果数及前 10 个结果中满足搜索意图的结果数。

对实验结果进行分析,发现相同搜索条件下,采用本文算法所获得的结果中,与用户兴趣相关的结果较传统算法排名更靠前。例如实验中的其中一次搜索,如表 3 所示。

表 3　排序结果对比

排名	传统个性化算法	本文算法
1	1	1
2	1	1
3	1	1
4	1	1
5	0	1
6	1	0
7	1	1
8	0	1
9	0	0
10	0	0

在表 3 中,1 表示满足用户搜索意图的结果;0 表示相关度不大的结果。可以看出,使用本文算法后满足用户搜索意图的结果增多了,且排名更加靠前,便于用户查找。然后将实验结果用查准率来评价比较,对多次结果进行综合分析取均值,各关键词的查询结果如表 4 所示。

表 4　查准率对比　%

查询条件	传统算法	本文算法	对比提升
气候变化	23	23	0
数据挖掘	55	70	15
Web 服务	56	84	18
定位技术	20	21	1
无线传感器	30	40	10
访问控制	50	62	12
文本分类	75	90	15
云存储	72	85	13
排序算法	31	42	11
大数据	61	86	25
模糊检测	20	22	2
图像处理	31	34	3
模式识别	45	49	4
用户兴趣	72	78	6
滤波	60	65	5

为了更直观地观察应用 2 种算法带来的排序质量,将对比结果用折线表示,如图 4 所示。

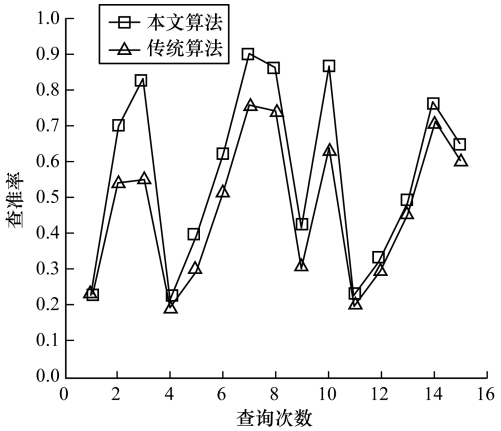

图 4　传统个性化排序算法与本文算法查准率的比较

由图 4 可知,在同一搜索环境下进行相同搜索时,采用本文算法的排序结果优于传统的个性化排序算法,更加适应对用户兴趣的判断,模拟用户真实的兴趣变化。因此,本文提出的基于兴趣衰减的个性化排序算法改善了传统个性化排序算法的不足,提高了个性化排序算法的性能,进而提高了搜索引擎的查准率。

5　结束语

本文考虑兴趣向量模型及人类遗忘特征,提出一种基于兴趣衰减变化的个性化排序算法。实验结果表明,该算法增强了用户兴趣判断的准确性,将符合用户搜索意图的结果排名提高,在搜索准确性方面有较大提升。由于最小遗忘时间间隔、最大遗忘时间间隔及衰减速率都为实验值,而实际上,不同人的记忆及兴趣周期都是有区别的,下一步将对该问题做深入研究。

参考文献

[1] 常　璐,夏祖奇. 搜索引擎的几种常用排序算法[J]. 图书情报工作,2003(6):70-73.

[2] XING W, GHORBANI A. Weighted PageRank Algorithm[C]// Proceedings of Annual Conference on IEEE Computer Society. Washington D. C., USA: IEEE Press, 2004: 305-314.

[3] PAULSELVAN M, SEKAR A C, DHARSHINI A P, et al. Survey on Web Page Ranking Algorithms[J]. Foundation of Computer Science, 2012, 41(19): 1-7.

[4] PAGE L, BRIN S, MOTWANI R, et al. The PageRank Citation Ranking: Bringing Order to the Web[J]. Stanford Digital Libraries Working Paper, 1998, 9(1): 1-14.

[5] DUHAN N, SHARMA A K, BHATIA K K. Page Ranking Algorithms: A Survey [C]//Proceedings of Advance Computing Conference. Washington D. C., USA: IEEE Press, 2009: 1530-1537.

(下转第 227 页)

5　结束语

本文基于免疫系统的识别、记忆及学习能力,提出了一种环境识别免疫算法处理动态环境背包问题,采用贪婪修补方法提高抗体群的可行度,通过环境识别算子以及对相似环境初始群的产生方式加速算法跟踪相似环境的速度。将算法用于 2 种高维的动态背包问题,并与已有的动态环境算法进行比较,仿真结果表明提出的算法具有解决高维组合优化问题的优越性,跟踪环境的速度较快。然而本文算法的一些参数需要结合经验及多次实验获得。因此,下一步的研究是减少参数或设计自适应调节策略,并在 TSP、投资组合等方面进行组合优化应用。

参考文献

[1] ZHANG Weiwei, YEN G G, HE Zhongshi. Constrained Optimization via Artificial Immune System[J]. IEEE Transactions on Cybernetics, 2014, 44(2): 185-198.

[2] LANARRIDIS A, STAFYLOPATIS A. An Artificial Immune Network for Multiobjective Optimization Problems[J]. Engineering Optimization, 2014, 46(8): 1008-1031.

[3] ALONSO F R, OLIVEIRA D Q, ZAMBRONI S A C. Artificial Immune Systems Optimization Approach for Multiobjective Distribution System Reconfiguration[J]. IEEE Transactions on Power Systems, 2015, 30(2): 840-847.

[4] 李枝勇,马　良,张惠珍. 求解 0/1 背包问题的自适应元胞粒子群算法[J]. 计算机工程, 2014, 40(10): 198-203.

[5] 汪定伟,王大志,王洪峰. 求解动态背包问题的多智能体进化算法[J]. 东北大学学报(自然科学版), 2009, 30(7): 948-951.

[6] 李志杰,李元香. 求解动态优化问题的多种群热力学遗传算法[J]. 计算机科学与探索, 2014(2): 179-185.

[7] 钱淑渠,武慧虹,涂　歆. 动态免疫优化算法及其在背包问题中的应用[J]. 计算机工程, 2011, 37(20): 216-218.

[8] 贺毅朝,宋建民,张敬敏,等. 利用遗传算法求解静态与动态背包问题的研究[J]. 计算机应用研究, 2015, 32(4): 1011-1015.

[9] ARAGÓN V S, ESQUIVEL S C, COELLO C A. Artificial Immune System for Solving Dynamic Constrained Optimization Problems[J]. Inteligencia Artificial, 2013, 14(4): 3-16.

[10] NASR G, HASSAN A. Multiobjective Genetic Programming for Financial Portfolio Management in Dynamic Environments[D]. London, UK: University College London, 2010.

[11] 郭小燕,王联国,代永强. 基于分段混合蛙跳算法的旅行商问题求解[J]. 计算机工程, 2014, 40(1): 191-194.

[12] NGUYEN T T, YANG Shengxiang, BRANKE J. Evolutionary Dynamic Optimization: A Survey of the State of the Art[J]. Swarm & Evolutionary Computation, 2012, 6: 1-24.

[13] CRUZ C, GONZÁLEZ J R, Pelta D A. Optimization in Dynamic Environments: A Survey on Problems, Methods and Measures[J]. Soft Computing, 2011, 15(7): 1427-1448.

[14] SIMÕES A, COSTA E. Improving the Genetic Algorithm's Performance when Using Transformation[M]// KURKOVA V, STEELE N C, NERUDA R. Artificial Neural Nets and Genetic Algorithms. Berlin, Germany: Springer, 2003: 175-181.

[15] SIMÕES A, COSTA E. An Immune System-based Genetic Algorithm to Deal with Dynamic Environments: Diversity and Memory[M]//KURKOVA V, STEELE N C, NERUDA R. Artificial Neural Nets and Genetic Algorithms. Berlin, Germany: Springer, 2003: 168-174.

[16] 鲁江林,何中市,陈自郁. 一种求解动态背包问题的离散粒子群优化算法[J]. 计算机科学, 2012, 39(9): 215-219.

[17] 贺毅朝,王熙照,寇应展. 一种具有混合编码的二进制差分演化算法[J]. 计算机研究与发展, 2007, 44(9): 1476-1484.

[18] 武慧虹,钱淑渠,徐志丹. 克隆选择免疫遗传算法对高维 0/1 背包问题应用[J]. 计算机应用, 2013, 33(3): 845-848.

[19] 王联国,洪　毅. 基于冯·诺依曼邻域结构的人工鱼群算法[J]. 控制理论与应用, 2010, 27(6): 775-780.

编辑　顾逸斐

(上接第 219 页)

[6] 李树青,韩忠愿. 个性化搜索引擎原理与技术[M]. 北京:科学出版社, 2008.

[7] 江　婕,李建民,曾勋炜. 基于用户反馈的个性化搜索引擎的研究[J]. 计算机与现代化, 2010(6): 116-118.

[8] 肖　瑜,赵俊忠. 一个新的个性化搜索引擎排序算法[J]. 太原科技大学学报, 2013, 34(3): 175-180.

[9] 文振威,秦　晓. 个性化搜索引擎的研究与设计[J]. 计算机工程与设计, 2009, 30(2): 342-344.

[10] KUMAR R, SHARAN A. Personalized Web Search Using Browsing History and Domain Knowledge[C]// Proceedings of Issues and Challenges in Intelligent Computing Techniques. Washington D. C., USA: IEEE Press, 2014: 493-497.

[11] 林　国,李伟超. 个性化搜索引擎中用户兴趣模型研究[J]. 软件导刊, 2012, 11(8): 26-28.

[12] 王洪伟,邹　莉. 考虑长期与短期兴趣因素的用户偏好建模[J]. 同济大学学报(自然科学版), 2013, 41(6): 953-960.

[13] 印桂生,崔晓辉,马志强. 遗忘曲线的协同过滤推荐模型[J]. 哈尔滨工程大学学报, 2012, 33(1): 85-90.

[14] 曾东红,汪　涛,严水发,等. 一种基于指数遗忘函数的协同过滤算法[J]. 科技广场, 2013(7): 10-15.

[15] 于　洪,李转运. 基于遗忘曲线的协同过滤推荐算法[J]. 南京大学学报(自然科学版), 2010, 46(5): 520-527.

[16] 张　磊. 基于遗忘曲线的协同过滤研究[J]. 电脑知识与技术(学术交流), 2014, 10(12): 2757-2762.

编辑　刘　冰