

文章编号:1671-6833(2017)06-0001-05

基于可能性条件偏好网络的交互式遗传算法及其应用

孙晓燕,朱利霞,陈 杨

(中国矿业大学 信息与控制工程学院,江苏 徐州 221008)

摘 要:根据用户实施的人机交互行为而隐式地获取用户偏好的交互式进化优化算法,可有效减轻用户疲劳,提高个性化搜索或推荐的效率.但是,已有研究没有考虑用户交互行为和偏好的不确定性,影响了对用户偏好的拟合精度以及基于该偏好表达的进化搜索.针对该问题,提出基于可能性条件偏好网络的交互式遗传算法,来刻画用户交互行为和偏好的不确定性,并提高算法的搜索性能.首先,采用交互时间表示交互行为,考虑交互行为的不确定性,给出交互时间可信度的定义,并基于该定义给出了用户不确定偏好的表达函数;其次,利用可信交互时间和偏好函数,定义了用户对评价对象的偏好权重,并利用该权重,设计(更新)可定量表示用户不确定偏好的可能性条件偏好网络,以更好地拟合用户偏好;然后,结合评价不确定性和可能性条件偏好网络,提出了改进的个体适应值估计策略,以更好地引导搜索;最后,将所提算法应用于图书个性化搜索中,结果表明了算法搜索的可靠性和高效性.

关键词:交互式遗传算法;不确定性;可能性条件偏好网络;个性化搜索

中图分类号:TP181 文献标志码:A doi:10.13705/j.issn.1671-6833.2017.06.001

0 引言

在当前信息社会中,个性化搜索、个性化服务等已成为网络发展、数据管理和挖掘的首要任务.实质上,个性化信息获取是优化问题.但是,目前从优化角度解决该类问题的研究成果还较少.在个性化信息获取中,人-机交互贯穿整个过程,如用户兴趣建模和跟踪等<sup>[1]</sup>.若将用户兴趣模型与进化优化算法结合,则可望提高用户搜索效率,而这恰好符合交互式进化优化算法应用的范畴<sup>[2]</sup>.交互式进化优化算法将人的智能与进化优化搜索过程进行结合,可有效解决、个性化产品设计、电力系统优化、房屋装修、网页布局、音乐创作、图像检索等问题<sup>[3-8]</sup>.

孙晓燕等<sup>[9]</sup>借鉴个性化搜索中用户兴趣建模的研究成果,考虑用户搜索的评价对象属性间的相互关联关系,提出了基于交互行为和条件偏好网络的交互式进化优化算法,并用于解决个性化搜索问题.然而,在该算法框架中,没有考虑用户认知或偏好评价的不确定性.笔者在文献[9]的基础上,研究含认知不确定性的隐式评价交互式进化优化算法.

由于鼠标滚动信息也可反映用户的偏好<sup>[10]</sup>,因此,可利用浏览时间和鼠标滚动次数定义交互时间和评价偏好的不确定性函数.因条件偏好网络(conditional preference networks, CP-nets)无法反映节点以及节点之间存在的偏好不确定性,文献[11]提出了一种新的偏好网络——可能性网络(probabilistic networks),它能灵活地表达不确定偏好.在此基础上,Cornelio<sup>[12]</sup>和 Bigot 等<sup>[13]</sup>考虑动态不确定和噪音特性,进一步构建动态概率 CP-nets,即每个依赖关系都以一定的概率存在,从而灵活刻画偏好关系.笔者在此基础上利用 PCP-nets (probabilistic conditional preference networks)可灵活表示不确定偏好.

1 基于交互行为的用户偏好不确定性表示

在电子商务搜索系统中,用户的交互行为能很好地体现用户的兴趣偏好.若仅仅利用交互时间的长短去刻画用户对某搜索对象的偏好度难以反映用户交互行为中的所有信息.如用户对第  $i$  个商品  $X_i$  和第  $j$  个商品  $X_j$  的交互浏览时间关系为  $t(X_i) > t(X_j)$ ,并不能说明用户对  $X_i$  的兴趣度一

定比对  $X_j$  的兴趣度大,有可能是  $X_i$  的页面较长,增加了用户的浏览时间,那么  $t(X_i)$  和  $t(X_j)$  就有一定的可信度. 所以,笔者结合鼠标滚动次数,考虑具有可信度的交互行为.

### 1.1 交互行为的可信度

考虑 4 类交互行为:①点击浏览行为  $A_1$ ;②保存行为  $A_2$ ;③收藏/标签行为  $A_3$ ;④决策/购买行为  $A_4$ . 偏好关系为: $A_1 < A_2 < A_3 < A_4$ .

采用交互时间刻画上述 4 种交互行为,并利用鼠标滚动次数定义交互时间的可信度. 对搜索对象  $X_i$ ,设用户交互时间为  $t_j(X_i)$ ,鼠标滚动次数为  $n_j(X_i)$ , $j = 1, 2, 3, 4$ ;  $i = 1, 2, \dots, L_j$ ,其中  $L_j$  表示对象集合包含的元素个数. 融合可信度的交互行为量化表达如式(1)所示:

$$t'_j(X_i) = t_j(X_i) e^{-\frac{n_j(X_i)}{\beta}}, \quad (1)$$

式中,  $\beta$  为控制条件可信程度变化的参数.

### 1.2 含交互不确定性的偏好表示

基于上述交互行为可信度表示,进一步给出用户对搜索对象  $X_i$  的偏好表示. 用户对于特别喜欢和特别不喜欢的搜索对象往往给出的评价越可信;而对于不关心的对象,评价的可信性则相对较小. 那么用户偏好不确定度定义如下所示:

$$g_j(X_i) = e^{-\frac{[g_j(X_i) - t_0]^2}{\gamma}}, \quad (2)$$

式中,  $\gamma$  为可变参数;  $t_0$  为平均浏览时间.

$$t_0 = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^4 t'_j(X_i). \quad (3)$$

当  $t'_j(X_i) < t_0$  时,说明用户对商品不感兴趣;当  $t'_j(X_i) \geq t_0$  时,说明用户对商品感兴趣,所以,给出用户对  $X_i$  的偏好函数如式(4)所示:

$$f_j(X_i) = \begin{cases} k e^{-g_j(X_i)}, & t'_j(X_i) \geq t_0; \\ k e^{-1} g_j(X_i), & t'_j(X_i) < t_0, \end{cases} \quad (4)$$

式中,  $k$  为调节参数.

用户对搜索对象  $X_i$  的偏好比重  $w'_{ji}$ ,  $i \in \{1, 2, \dots, L_j\}$ ,  $j \in \{1, 2, 3, 4\}$ , 如式(5)所示:

$$w'_{ji} = t'_j(X_i) f_j(X_i). \quad (5)$$

对式(5)进行归一化处理:

$$w_{ji} = \frac{w'_{ji}}{\sum_{j=1}^4 \sum_{i=1}^{L_j} w'_{ji}}. \quad (6)$$

## 2 基于不确定偏好权重的可能性条件偏好网络的构建

### 2.1 可能性条件偏好网络的初始化构建

采用文献[9]的方法构建初始化条件偏好有

向图,不再赘述,重点是可能性条件偏好表的构造. 设 PCP-nets 的偏好有向图共包含多层节点,其中第  $l$  层有  $m_l$  个节点变量,记为  $x_1^l, x_2^l, \dots, x_{m_l}^l$ ,第  $k$  个变量  $x_k^l$  具有  $n_k$  个取值  $x_{k,1}^l, x_{k,2}^l, \dots, x_{k,n_k}^l$ ,那么  $x_{k,j}^l$  在所有父节点可能取值组合条件下的偏好可能性设共有  $M_{kj}^l$  种组合,各偏好可能性取值记为  $\alpha_{k,j,1}^l, \alpha_{k,j,2}^l, \dots, \alpha_{k,j,M_{kj}^l}^l$ . 对于父节点具有明确偏好属性取值的  $\alpha_{k,j,p}^l = 1$ ,其余的  $\alpha_{k,j,p}^l$  在  $[0, 1)$  随机取值.

### 2.2 可能性条件偏好网络的更新

首先,可能性条件偏好表的更新. 通过式(4)可获得决策变量的偏好度. 设决策变量的第  $m$  个取值  $x_{l,m}$  在  $A_j$  交互行为下出现次数为  $n_j^{lm}$ ,那么  $x_{l,m}$  的偏好度如式(7)所示:

$$f(x_{l,m}) = \frac{\sum_{j=1}^4 f_j(x_l) n_j^{lm}}{\sum_{j=1}^4 n_j^{lm}}. \quad (7)$$

其次,条件偏好有向图的更新. 由式(6)可得用户在交互行为  $A_j$  下对评价对象  $X_i$  的偏好权重为  $w_{ji}$ . 设当前搜索群体中,决策变量  $x_l$  在  $A_j$  交互行为下的取值共有  $n_j^l$  个,决策变量的偏好权重如式(8)所示:

$$w(x_l) = \frac{\sum_{j=1}^4 w_j n_j^l}{\sum_{j=1}^4 n_j^l}. \quad (8)$$

根据式(8)计算偏好权重,更改支配关系.

## 3 基于可能性条件偏好网络的交互式遗传算法

### 3.1 基于可能性条件偏好网络的个体适应值的估计

首先,考虑条件偏好有向图中决策变量的层级分布贡献度.  $x_j$  支配的决策变量越多,  $class(x_j)$  越小. 根据文献[9]可知,决策变量  $x_j$  的贡献度为  $C_1(x_j) = 2^{(class(x_j)-1)}$   $class(x_j)$ ,若某决策变量未出现,则设该贡献度值为  $C_1(x_j) = 1$ . 结合偏好可能度,借鉴贝叶斯推理的链式公式,对于进化个体  $X_i = \{x_{i1}, x_{i2}, \dots, x_{in}\}$  的适应值估计策略如式(9):

$$F(X_i) = \prod_{j=1}^n 2^{(class(x_j)-1)} class(x_j) \alpha(x_{j,i_j}^{class(x_j)}). \quad (9)$$

### 3.2 算法步骤

Step1: 根据实际问题进行编码,并设定遗传操作的各种参数;

Step2:基于用户输入关键词信息,历史搜索信息和社会群体搜索结果,获得初始化种群;

Step3:根据 2.1 节方法,构建初始化可能性条件偏好网络;

Step4:根据 3.1 节方法,估计当前进化种群的个体适应值,选择  $N$  个适应值较大的进化个体提交给用户;

Step5:用户实施交互行为,若找到满意解,则算法终止,并输出满意解,否则,转 Step6;

Step6:记录浏览时间和鼠标滑动次数,根据第 1 节计算用户对评价对象的偏好权重,并根据 2.2 节方法更新可能性条件偏好网络;

Step7:对当前种群实施选择,交叉和变异操作,生成新种群,转 Step4.

## 4 算法的应用

### 4.1 实验背景

将本文算法应用到文献[9]所构建的图书个性化搜索系统中,如图 1 所示.通过算法的比较,验证本文算法的有效性.为了说明本文算法 IGA-PCP(probabilistic conditional preference network assisted interactive genetic algorithm)的整体性能,将其与文献[9]所提算法 IGA-CP(conditional preference network assisted interactive genetic algorithm)以及传统的交互式遗传算法 IGA(interactive genetic algorithm)进行比较.



图 1 图书搜索系统界面

Fig.1 Interface of book search system

### 4.2 参数设置

采用二进制形式编码,编码串按照决策属性分块.以当当网心理学图书为例,将心理学图书属性分为 7 类,每一类属性下有若干取值,本实验中共 46 个,需要 19 位二进制码.实验中,  $\beta = 6$ ,  $k = 5$ ,  $\gamma = 4$ , 3 种比较算法均采用轮盘赌选择,单点交叉和变异,且交叉和变异概率分别为 0.6、0.1,种群规模设为 8.

### 4.3 实验结果与分析

#### 4.3.1 偏好不确定函数表示的合理性

鼠标滚动次数以及可信时间之间的关系如表 1 所示.可信时间与所定义的偏好度函数关系如图 2 所示.

从表 1 可知,鼠标滚动次数与用户实际交互时间的相互关系与本文所提可信时间相符,表明笔者所提时间可信度的合理性.由图 2 可知,式(4)所定义的偏好度函数随着可信时间的增加而非线性增加,符合实际情况,表明了偏好度函数定义的合理性.

表 1 IGA-PCP 用户浏览行为			
Tab. 1 IGA-PCP user browsing behavior			
图书个体	用户实际交互时间/s	鼠标滚动次数	可信时间/s
儿童发展与教育心理学	15	2	11
自闭症	18	4	9
学前儿童心理学	20	6	7
特殊儿童及青少年心理学	50	8	13
魔鬼心理学	4	5	2
重口味心理学	16	4	8
女人气场心理学	1	1	1
女性心理成长自疗课	20	9	4

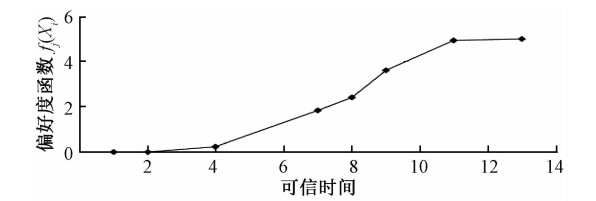


图 2 偏好度函数  $f_j(X_i)$  与可信时间的关系

Fig.2 Relation between preference function  $f_j(X_i)$  and credible time

#### 4.3.2 对比实验

3 种比较算法,实验结果如图 3 所示.实验针对 84 类图书共 4 420 本,在所有的图书中找到“特殊儿童的问题行为干预”,记录实验数据如图 3 所示:①IGA-PCP 平均互异个体的比率最高,这说明了本文算法 IGA-PCP 能有效提高搜索的多样性.②本文算法的搜索时间比传统 IGA 大约少一半,并且明显小于 IGA-CP 算法搜索时间.③IGA-PCP 用户实施交互操作行为的次数平均为 4.0,IGA-CP 平均为 7.6,传统的 IGA 平均为 7.7,由此可以看出,本文算法 IGA-PCP 可有效减轻用户疲劳.

为了说明所提算法的整体性能,现对励志类图书进行搜索,实验针对 24 类图书共 1 159 本,实验结果如表 2 所示.从表 2 可以看出,本文算法不仅有效减轻用户疲劳,而且可有效提高算法多

样性. 总体来说,本文算法明显优于其他两种算法.

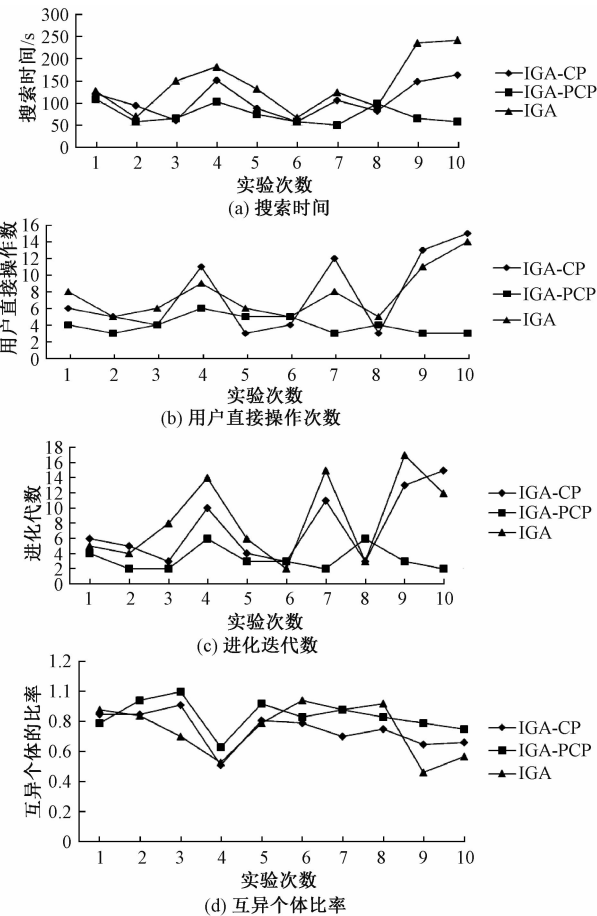


图3 84类心理学图书搜索实验结果

Fig.3 The results of the experiment of eighty four kinds of psychology books

表2 24类励志图书搜索实验结果

Tab.2 The results of the experiment of twenty-four kinds of inspirtional books

算法	平均进化代数	平均搜索时间/s	用户交互次数	互异个体比率
IGA-PCP	5.3	93.4	5.9	0.848
IGA-CP	7.8	112.6	7.8	0.689
IGA	9.6	126.0	7.2	0.669

5 结束语

用户在搜索过程中,用户的交互行为能够有效地反映用户偏好信息,但由于用户认知的不确定性,反映到交互行为上也具有不确定性. 针对此,笔者考虑了交互行为的不确定性刻画,以及基于该不确定性刻画的用户不确定偏好的描述,在用户偏好建模方面,基于所考虑的不确定性偏好,提出了一种可能性条件偏好网络拟合用户评价偏好的方法,并给出了适应值估计策略,将算法应用

到图书搜索系统中,实验结果表明了该算法的可行性与有效性.

如何利用其他用户的信息构建多用户偏好模型,并进行有效集成和更新,以实现网络环境下的信息动态更新,提高搜索效率,将是下一步要研究的问题.

参考文献:

[1] QIAN X M, FENG H, ZHAO G S, et al. Personalized recommendation combining user interest and social circle [J]. IEEE transactions on knowledge data engineering, 2013, 26(7): 1763-1777.

[2] 孙晓燕. 高级交互式遗传算法理论与应用[M]. 北京: 科学出版社, 2010: 7-18.

[3] SORM D, RIMCHAROEN S. Web page template design using interactive genetic algorithm [C]//International Computer Science and Engineering Conference (ICSEC). Nakorn Pathom, Thailand: IEEE, 2013: 201-206.

[4] 杨胡萍, 李威仁, 左士伟, 等. 基于改进遗传算法的电力系统无功优化[J]. 郑州大学学报(工学版), 2015, 36(6): 66-69.

[5] LIU X. Color mapping design from image to 3D product model [J]. Journal of mechanical engineering, 2009, 45(10): 222-227.

[6] GARCIA-HERNANDEZ L, ARAUZO-AZOFRA A, SALAS-MORERA L, et al. Recycling plants layout design by means of an interactive genetic algorithm [J]. Intelligent automation and soft computing, 2013, 19(3): 457-468.

[7] KOGA S, INOUE T, FUKUMOTO M. A proposal for intervention by user in interactive genetic algorithm for creation of music melody [C]//International Conference on Biometrics and Kansei Engineering. Tokyo, Japan: IEEE, 2013: 129-132.

[8] DASS M V, ALI M M, ALI M R. Image retrieval using interactive genetic algorithm [C]//International Conference on Computational Science and Computational Intelligence. Las Vegas, USA: IEEE, 2014: 215-220.

[9] 孙晓燕, 陆宜娜, 巩敦卫, 等. 基于 CP-nets 的偏好感知交互式遗传算法及其个性化搜索[J]. 控制与决策, 2015, 30(7): 1153-1161.

[10] CLAYPOOL M, LE P, WASED M, et al. Implicit interest indicators [C]//Proceedings of the 6th International Conference on Intelligent User Interfaces. New York, USA: IEEE, 2001: 33-40.

[11] BENAMOR N, DUBOIS D, GOUIDER H, et al. Possibilistic networks: a new setting for modeling prefer-

ences[M]. Berlin, Germany: Springer, 2014:1 – 7.

[12] CORNELIO C, GOLDSMITH J, MATTEI N, et al. Dynamic probabilistic CP-nets [C]//Multidisciplinary Workshop on Advances in Preference Handling. Chicago, USA:IEEE, 2013:1 – 7.

[13] BIGOT D, FAGIER H, MENGIN J, et al. Probabilistic conditional preference networks[J]. Computer science, 2013,19:72 – 81.

Probabilistic Conditional Preference Network Assisted Interactive Genetic Algorithm and Its Application

SUN Xiaoyan, ZHU Lixia, CHEN Yang

(School of Information and Control Engineering, China University of Mining & Technology, Xuzhou 221008, China)

**Abstract:** Interactive evolutionary algorithms with user preference implicitly extracted from interactions of user were more powerful in alleviating user fatigue and improving the exploration in personalized search or recommendation. However, the uncertainties in user interactions and preferences have not been considered in the previous research, which might greatly impact the reliability of the extracted preference model, as well as the effective exploration of the evolution with that model. Therefore, an interactive genetic algorithm with probabilistic conditional preference networks (PCP-nets) was proposed, in which, the uncertainties were further figured out according to the interactions, and a PCP-net was designed to depict user preference model with higher accuracy by involving those uncertainties. First, the interaction time was adopted to mathematically describe the relationship between the interactions and user preference, and the reliability of the interaction time was further defined to reflect the interactive uncertainty. The preference function with evaluation uncertainty was established with the reliability of interaction time. Second, the preference weights on each interacted object were assigned on the basis of preference function and reliability. With these weights, the PCP-nets were designed and updated by involving the uncertainties into the preference model to improve the approximation. Third, a more accurate fitness function was delivered to assign fitness for the individuals. Last, the proposed algorithm was applied to a personalized book search and its superiority in exploration and feasibility was experimentally demonstrated.

**Key words:** interactive genetic algorithm; uncertainty; possibilistic conditional preference networks; personalized search