

电子商务技术中的智能自组织系统 及 CCM 部件设计^{*}

Intelligent Self-Organizing System in E-Commerce Technology and CCM Component Design

李海鹰 李 军 简 兵 庄镇泉

(中国科技大学电子科学与技术系 合肥 230026)

Abstract This intelligent self-organizing system aims at self-organizing and re-allotting Web-site' resource. It uses the characteristic clustering model (CCM) to mine the user's characteristic vectors and maps them to Web resource space to implement the Web resources' intelligent configuration. The kernel of the system is characteristic clustering model, which takes advantage of the ameliorated ART model to perform the user's characteristic vectors clustering. This paper introduces the system structure, and then describes the characteristic clustering model emphatically.

Keywords E-Commerce, Information capture, Vector-clustering, Derivative subsystem

1 引言

随着 Internet 技术的飞速发展,智能信息处理技术已经渗透到电子商务的各个应用领域。如何针对网络用户智能地提取用户典型行为特征,提供个性化服务,已经成为电子商务技术的一个重要研究方向。

智能自组织系统是针对个性化网站的资源自组织和资源再分配问题而设计的集成应用。个性聚类模型 (CCM) 是系统的核心,它由改进的二值模式 ART 算法实现个性矢量的聚类和个性矢量识别。智能自组织系统利用 CCM 挖掘网络用户典型行为的特征矢量,并通过该特征矢量映射到网站资源组织空间,从而实现网站的资源智能配置。该系统侧重整体建模,从全局的角度设计系统关键部件,规划系统的实现步骤。

典型用户行为特征代表了一个用户群体的行为特点。利用这种特征组织网络资源极大地减少了生成网站动态服务结构的计算量,同时将个性挖掘的计算和组织网页的计算在空间和时间上分离,客观上达到了负载均衡的目的。

本文首先总体介绍智能自组织系统的组成、功能和模块设计策略;其次,着重讨论作为智能自组织系统

核心的 CCM 算法;最后,描述了 CCM 部件在 C++ 语言环境下的实现过程。

2 智能自组织系统

智能自组织系统的关键问题是如何获得用户典型行为特征,系统的设计围绕这一中心进行。本系统功能概括为两种类型:用户典型特征的形成和个性化服务的推荐(典型特征的应用)。系统的工作过程如下:数据采集,捕获用户访问信息形成原始数据;数据预处理,分类重组用户访问信息,形成层次分类结构;数据聚类,提取特征矢量,并采用多维矢量聚类方法,形成同一类型用户行为的典型特征;数据跟踪,建立行为仿真模型,对典型特征矢量进行阶段性校正;数据应用,通过矢量分析,实现基于典型特征矢量空间模式的资源组织。

考虑到智能自组织系统是集成应用,可能会运行在不同的操作环境中,所以在系统设计中采用了模块集成的思想。智能自组织系统由六个模块构成,如图 1 所示。左边三个模块主要完成特征矢量的形成,在右边三个模块完成个性化推荐服务。下面逐一介绍这六个功能模块。

^{*} 本课题得到 973 国家重点基础研究发展规划项目(G1998030413)的资助。李海鹰 博士生,主要研究领域为网络与软件工程、神经网络、电子商务技术等,李 军 硕士生,主要研究领域为电子商务技术、网络系统工程、神经网络等,简 兵 硕士生,主要研究领域为电子商务技术、网络系统工程、神经网络等,庄镇泉 博士生导师,主要研究领域为神经网络、计算机工程和大规模集成电路设计等。

★**信息捕捉机** 能在 Web 服务器运行过程中动态地收集、发现信息,能使用策略对用户信息进行过滤和规划,信息捕捉机工作在数据采集阶段,它通过与数据库的通信,存贮和交换用户相关信息,并可以相应地调整采集参数。

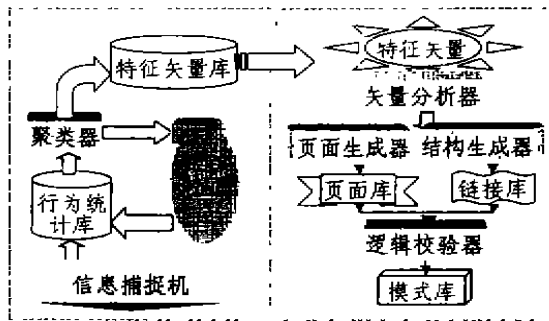


图1 智能自组织系统结构

★**聚类器** 智能自组织系统的核心部件之一,是在聚类模型的基础上设计的应用单元。聚类器可以有效挖掘网络用户行为典型个性特征,形成稳定的分类集,同时可以对个性特征矢量进行识别。聚类可以在不同层次进行,图2是电子商务商品交易网站的类型层次图。

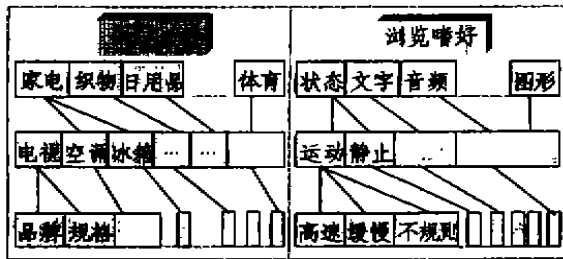


图2 个性特征分类层次图

对电子商务网站来说,个性特征可以通过两个方面体现出来:项目兴趣和浏览嗜好。项目兴趣和浏览嗜好特征结合形成个性矢量。

★**跟踪校正器** 调整聚类器参数的应用单元。它将用户个性特征矢量与用户原始信息进行比较,通过调整聚类算法弥补个性特征聚类中出现的明显缺陷。

★**矢量分析器** 利用特征矢量空间模式安排生成器作业流程的应用单元;本单元将特征矢量空间映射到网站的页面组织和结构组织空间上,指导网站生成器动态生成页面和链接结构。

★**生成器** 对应于分析器将用户典型特征矢量分解为项目兴趣和浏览嗜好两个方向,生成器相应地分为页面生成器和结构生成器。项目兴趣映射到结构生

成器、指导网页链接指针的组织,使用户尽快找到感兴趣的东西;浏览嗜好映射到页面生成器、指导页面的资源组织,使生成的界面更加适合用户的口味,分析器和生成器的设计质量关系到个性化网站自组织系统最终输出的效果和系统自组织的效率问题。

★**逻辑校验器** 对生成的个性化网页结构进行逻辑测试的应用单元;本单元具有检查页面组织的合理性以及链接结构的可行性能力。如果页面组织和链接组织通过检验,逻辑校验器正式将它们存入模式库,供个性检索系统调用。

智能自组织系统根据需要分布在不同计算环境中,依照系统运行的逻辑顺序依次完成控制程序指派的任务。主控系统依照策略对自组织系统进行管理,控制流程如图3所示。主控系统的控制有两个方向。第一,获取个性化信息,并保存此信息;第二,利用模式匹配,找到同类的已经存在的个性化信息,实现个性化推荐服务。

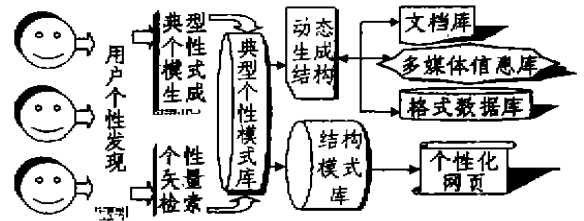


图3 主控系统控制流程

主控系统的设计策略包括:(1)数据发布策略:系统使用的数据主要包括:访问信息、个性表现数据、页面结构数据,数据分布策略通过设计数据库结构完成;(2)任务分布策略:分解系统任务,规定系统部件职能。任务的分配依据部件职能直接确定;(3)协作控制策略:规划系统流程,明确部件之间的逻辑依赖关系;分层次控制部件工作进程和控制信息传输顺序,消除部件冲突。

3 CCM 算法设计

个性聚类算法以二值 ART 算法为基础,引进了矢量匹配度量(VDM)概念计算矢量相似度,用模式匹配和角匹配结合的方法解决输入序列的问题。CCM 算法有两个过程:输入激活过程和失配搜索过程。

1)激活过程:输入矢量通过输入节点进入网络,并与各输出节点(又称处理单元 PE)所对应的参考矢量进行相似性比较,其中与输入模式最接近的处理单元被激活,该激活单元再将其对应的模板矢量与输入模式进行匹配并与事先给定的警戒阈值进行比较,若超过警戒阈值则发生谐振,激活的处理单元所对应的激

活参考矢量与匹配模板矢量均按照相应的权值调整规则向输入模式移动。

2) 失配搜索过程: 若匹配结果小于警戒阈值, 则不发生谐振, 激活的处理单元被废除, 重新选择新的激活单元; 如果已有的处理单元均不能通过警戒阈值的检测, 则产生一个新的处理单元来匹配当前的输入模式。

聚类算法如下:

```

for each(p=1..m)
  //m 为已知类别数, p 为匹配次数控制参数
  输入二值矢量 X;
   $K_j = \sum_{i=1}^n B_{ji} X_i$ ; // 计算输出节点的加权值
  ////////////////
  for (each j=1..m)
    {  $K_j = \max_{i \in \{1, 2, \dots, m\}}$  { $K_j$ }; }
    // 节点之间的横向竞争, 最大值为取胜节点
  ////////////////
   $X' = T_j \cap X$ ; // 由节点 j 发出代表典型模式的权值信号
  //  $T_j$  与 X 进行与操作, 形成新的输入矢量
  // 矢量 X 和 X' 的相似度是通过计算矢量匹配度(VDM)
  // 来完成的。
   $MDM(X', X) = \min(|X'|, |X|)$ ; // MDM 为模匹配度
   $\cos\theta = (X' \cdot X) / (|X'| \cdot |X|)$ ; //  $\cdot$  为点乘
   $ADM(X', X) = (\pi - \theta) / \pi$ ; // ADM 为角匹配度
   $VDM(X', X) = MDM(X', X) * ADM(X', X)$ ;
  // VDM 为矢量匹配度
  ////////////////
  if ( $VDM(X, X') \geq \rho$ ) //  $\rho$  为阈值,  $\rho \in (0, 1)$ 
    {  $X \in_j$  类;
       $T_j(t+1) = T_j(t) * X$ ;
       $B_j(t+1) = T_j(t) / (0.5 + \sum T_j(t) * X)$ ; break; }
    else { 删除 j 节点; 输入信号复位; 继续匹配过程; }
  ////////////////
}
if ( $p == m$ ) { 增加一个新的类节点; } // 如果搜索的整个过程
// 都达不到相似度的要求, 便按照输入信号生成新类, 赋
// 初始权值。

```

算法中变量说明: X: 二值矢量输入信号; Y: 输出, 代表已经存在的模式; B_{ji} : 输入节点 i 到输出节点 j 的权; T_{ji} : 输出节点 j 到输入节点 i 的权。

4 CCM 部件设计

CCM 建立在改进的 ART 模型的基础上, 在 C++ 语言环境中实现。CCM 部件设计包括以下几个方面: 神经元设计, 神经元之间协同工作的控制策略。模型采用类封装结构, 由两类神经元组成, 每个神经元均是一个具有独立处理能力的智能实体。下面具体介绍 CCM 的部件设计。

4.1 神经元设计

4.1.1 神经元基类 CCM 的两类神经元在输入输出方式、数据存储、基础功能函数等方面有一定的共性, 通过对这些共性的分析, 对应 C++ 中类的设计层次结构, 可以在抽取公共模式的基础上构造神经元基类。神经元基类可以派生出模型所需的执行各种功能的神经元。聚类神经元基类(CCellBase)组成如图 4 所

示。

数据类型定义: 公共数据结构的类型说明, 例如, 权值向量类型;

公共函数定义: 定制特殊的数学处理公式;

预处理函数定义: 负责对数据进行预处理, 例如, 数据初始化函数;

数据基本 I/O 定义: 负责说明数据的对外通讯方式, 主要功能是读/写外部数据文件;

继承类型定义: 继承其它基础类的功能, 可扩展类神经元基类功能。

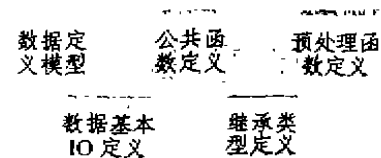


图 4 神经元基类(CCellBase)组成

4.1.2 注意子系统神经元类函数(CAdvertCell)

注意子系统神经元是 CCM 的入口, 接受三种信号输入: 原始分类数据, 增益控制和反馈(外星信号)。原始分类数据来自于信息捕捉机, 是一个二值(0,1)输入信号矢量; 增益控制信号和反馈信号均来自于取向子系统, 其中增益控制信号是控制状态信号(0,1), 反馈信号是分类模式信号。注意子系统神经元的功能是通过 2/3 规则函数处理三种输入信号, 判别神经元状态, 如果神经元处在兴奋状态会产生输出, 向上层神经元运动; 如果神经元呈抑制状态, 则继续等待接受数据。

```

class CAdvertCell: public CCellBase
{protected:
  int input;
  int output;
  int plus;
public:
  BOOL Regulation(int X, int G, int P); // 2/3 规则函数, X
  // 输入, G 增益, P 外星信号
  int Transmit(BOOL R, int X); // 输出信号函数
  void Adjust(int X, int P, int T) // 调整函数
  BOOL Status(BOOL R, int X, int G, int P, BOOL A);
  // 神经元状态函数
}

```

4.1.3 取向子系统神经元类函数(CTropismCell)

取向子系统神经元处理来自于注意子系统的刺激信号, 并求出加权值, 供同类神经元产生横向竞争使用。在竞争中胜利的神经元被激活, 向注意子系统输出分类模式信号(外星信号)和增益控制信号。神经元类的构成:

```

class CTropismCell: public CCellBase
{protected:
  int input;
  int output;
  float temp;
public:

```

```
float Warfare(int X,float *W); //产生竞争加权数值
int *ExpModel(int Y); //输出信号函数
BOOL Status(int Y); //神经元状态函数
```

4.2 两类子系统协同工作的控制策略

注意子系统和取向子系统既是两个相对独立的智能体,又是需要密切协同工作的部件.两类子系统的协同工作关系由控制策略来定义,控制策略的内容包括:
a)协作请求策略,该策略解决数据接收或发送的控制;
b)复位请求策略,该策略解决数据重新输入的控制;
c)产生竞争策略,该策略解决神经元状态激活的控制;
d)增加神经元请求策略,该策略解决新类的产生控制。

如图5所示,由于这些控制策略的存在,可以消除神经元数据滞留、输入/输出时间序列偏差、无意义竞争等方面的问题,达到子系统协调工作的目的。

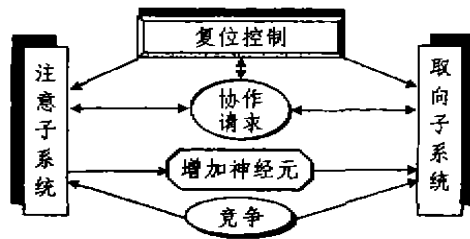


图5 控制策略示意图

4.3 逻辑运行过程

基于以上设计的一个完整的CCM部件及控制策略的方案如图6所示,它以数据流动方向为线索来说明,数据的流动是通过函数之间互相调用实现的。模型的主要函数存在于神经元基类和衍生出来的两个子类CAdvertCell和CTropismCell之中,父类和子类之间的继承关系,使函数之间可以顺利衔接。

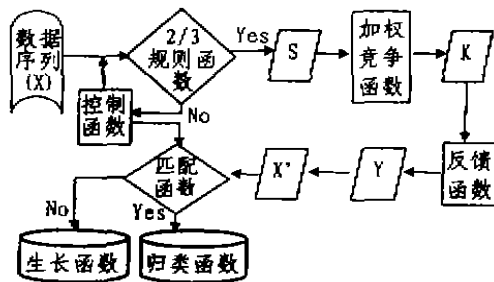


图6 数据流动和函数关系

模型中流动的数据有以下几类:原始数据X;注意神经元的输出信号S;加权矢量K;优势分类模式Y;

取向神经元的输出新模式X'。图6的具体说明如下:X在控制函数的作用下,进入注意神经元部件;经过注意神经元的2/3规则函数判断,产生S;在协作请求策略的控制下,S受到加权函数的推动,进入取向神经元,同时竞争策略发挥作用,展开同类竞争;竞争函数最终产生一个Y;Y在协作请求策略作用下,返回注意神经元,同时形成X';注意神经元中的匹配函数对X和X'进行分析,得到矢量匹配度(VDM);VDM与预置的警戒阈值比较,决定调用归类函数进行分类,还是执行增加新神经元策略,调用生长函数生成新模式。

结束语 个性化服务是电子商务网站提高网络效率和吸引网络用户访问的有效方法,商务网站的资源组织方式、组织效率已经成为智能电子商务技术的焦点之一,智能自组织系统的优点在于采用面向对象的思想设计模块架构,使模型的脉络清晰,移植性强,可读性更好。

智能自组织系统提出了典型用户行为特征的概念,并利用此特征矢量较为成功地解决了个性化网站建设中遇到的兴趣挖掘、服务推荐和动态网页生成等个性化服务中的关键性问题。该智能自组织系统可以针对不同用户行为,自主地分析用户的需求,并加以监测和校正,为用户提供个性化的服务。

参考文献

- 1 庄镇泉,王煦法.神经网络和神经计算机.中国科学出版社,1992.12
- 2 焦李成.神经网络计算.西安电子科技大学出版社,1993.9
- 3 Bennett D. Visual C++5 Developer's Guide
- 4 Carpenter G A. The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network Trans. IEEE on Computer, 1988(March), 77~88
- 5 Sanvhez J A, Legget J H. Agent services for users of digital libraries. J of Network and Computer Applications, 1997, 20: 45~58
- 6 Sycara K, et al. Distributed intelligent Agents. IEEE EXPERT, 1996, 11(5): 36~46
- 7 Joachums T, Freitag D, Mitchell T. WebWatcher: A Tour Guide for the World Wide Web. In: Proc. of the 15th Intl Joint Conf. on Artificial Intelligence IJCAI-97. 1997. 770~775
- 8 Armstrong R, Freitag D, Joachums T, Mitchell T. Web-Watcher: A learning apprentice for the world wide web. In Working Notes of the AAAI Spring Symposium: Information Gathering from Heterogeneous, Distributed Environments, Stanford University, AAAI Press, 1995. 6~12