免疫进化机制及其在时序模式挖掘中的应用研究

杨炳儒[1]，秦奕青[1,2]，宋泽锋[1]

（1. 北京科技大学信息工程学院，北京 100083；2. 北京机械工业学院计算机与自动化系，北京 100085）

[摘要] 针对目前动态数据挖掘中存在的问题，提出基于数据增量的动态挖掘进程概念；在动态挖掘进程和生物免疫进化过程的相似性基础上，提出了知识发现中的免疫进化机制的基本内涵；给出了基于免疫进化机制的时序模式挖掘算法及其实验分析，以验证理论的正确性和有效性。

[关键词] 动态数据挖掘；免疫算法；动态挖掘进程；免疫进化机制；时序模式挖掘

[中图分类号] TP311 [文献标识码] A [文章编号] 1009－1742(2008)04－0084－06

1 前言

许多知识发现系统需要从规律性变化的数据中准确挖掘信息。在这样的系统中，无论是频繁的还是偶然的数据更新都有可能改变先前发现的知识或规则的状态。因此，知识的维护和更新就成为知识发现研究中要解决的问题之一。

目前，知识发现研究采用增量更新、数据快照和时间戳等方法对知识进行动态维护和及时更新。以关联规则为例，D. W. Cheung 等研究了一系列维护更新问题[1]；何炎祥等在此基础上提出应用抽样技术的算法 SEA[2]；冯玉才等提出 IUA/PIUA[3] 算法；宋余庆、杨明、吉林根等先后提出了几种关联规则更新算法[4~6]。这些算法基于动态库和交互方式，以数据量增加后规则的维护算法居多，这样的动态挖掘算法一般称为增量式挖掘算法。一般来说，增量式数据挖掘的研究是以静态挖掘为基础的。如上述算法其本质都是以 Apriori 算法[7] 为基础，考虑数据库的动态性或者约束条件的改变，以有效减少数据库扫描次数和候选项目集为目标，实现增量式关联规则的更新维护，得到比较精确的结果。

但是这类算法存在几方面问题。一方面，在结果分析上，由于挖掘是在新的时间断面上进行，所以不可避免会造成一次数据中随机因素的增大，对生成或者丢弃的规则和知识产生扰动。另一方面，由于大型数据库中海量数据的出现，使得原来的数据挖掘算法存在实现上的困难。再一方面，虽然在数据库的数据积累过程中，知识库的结构具有一定的稳定性，但它也在随着数据的积累不断变化。因此，有必要整体考察知识库中的知识随时间动态变化的特点和规律，不再简单地以一次挖掘结果来决定知识的取舍。

根据数据库和知识库动态变化特性，引入免疫进化方法以解决上述问题，因为免疫进化计算适合于直接解决动态问题，其核心思想就是利用进化历史中获得的信息指导搜索或计算。

2 基于数据增量的动态挖掘进程概念

简单地说，动态数据挖掘是“运动的”挖掘，即在静态挖掘的基础上，动态地考察每次挖掘的结果，综合给出对知识的评价。实际上，动态数据挖掘就是从知识的演化角度，综合评价知识的类型，给出对知识的取舍。它要求知识发现系统不仅要考虑基于动态数据库的增量式挖掘，而且要考虑基于动态知识库的动态挖掘算法。

在动态数据挖掘中，作为挖掘结果的知识主要

[收稿日期] 2007－02－06；修回日期 2007－05－28

[基金项目] 国家自然科学基金资助项目(60675030)；国家科技成果重点推广计划资助项目(2003EC000001)

[作者简介] 杨炳儒(1943－)，男，天津市人，北京科技大学教授、博士生导师，主要研究方向为知识发现与智能系统，柔性建模与集成技术等；秦奕青(1969－)，女，北京市人，北京机械工业学院副教授，北京科技大学博士研究生，主要研究方向为数据挖掘

受以下几个方面的环境因素影响:a. 动态变化的数据库;b. 动态变化的知识库;c. 动态变化的用户感兴趣度;d. 挖掘相关领域的时空变迁;e. 论域与阈值的变化等,其中最重要的是时变数据、背景知识和用户兴趣。

定义1 动态数据挖掘的环境空间是5元组 $\Omega = (U, V, R, S, T)$,其中,U 表示动态数据库,V 表示动态知识库,R 表示用户感兴趣度论域;S 表示变迁的时空域;T 表示论域与阈值集。

对于时序数据库,数据会随着一定的时间间隔不断地加入数据库。在考虑其增量数据添加后知识发现中的动态挖掘问题时,首先规定一定的时间间隔,对挖掘的数据库 DB 采用如下的分库方案。

方案1 如果数据库的字段中包含时间属性 T,则根据 T_i 将数据库逻辑地划分成 n 个 DB_i,使每个 DB_i 与 T_i 对应。其中,$T_i = \{tractime_i, tractime_j\}$ 为一时间区间,且满足:如果 $i<j$,那么 $tractime_i < tractime_j$, $i, j = 1, 2, \cdots, n$。

方案2 如果数据库的字段中不包含时间属性,那么将数据库 DB 逻辑地分成 n 个 DB_i,满足当 $i<j$ 时,$DB_i < DB_j$,且 $DB_n = DB$。

在此分库方案的基础上,提出了动态挖掘进程的概念。

定义2 基于数据增量的动态挖掘,通过对数据库的划分方案对逻辑子库依据时间顺序不断挖掘知识,综合每次挖掘结果给出对知识的评价过程,称为动态挖掘进程。

动态挖掘进程与增量式挖掘既相互联系又有所区别。一方面,动态挖掘进程融合了增量式挖掘。它不仅依靠增量式挖掘方法实现当前的挖掘,而且要利用每次挖掘结果进行动态分析,剔除不确定因素对数据造成的影响,指导每次挖掘。另一方面,基于数据增量的动态挖掘进程不同于单纯的增量式挖掘。增量式挖掘注重具体算法的研究,在当前挖掘中尽量利用前一次挖掘的结果,使当前挖掘的算法复杂度降低,以求算法的高效性和有效性。而基于数据增量的动态挖掘进程不仅针对变化的数据库,在每次挖掘时选用具体的增量式挖掘算法,以求得一次挖掘的结果;而且它在对知识的评价运用上,不是依据前一次挖掘的结果,而是根据每一次挖掘结果,综合地历史地对挖掘的规则进行评价。在使用动态挖掘进程进行知识发现时,甚至可以不注重具体的挖掘结果,而通过对知识或者规则的跟踪性研究,使用户挖掘出自己感兴趣的知识。这种历史地、综合地跟踪性研究可以很好地避免一次挖掘中数据的随机因素对挖掘结果造成的干扰。

按照哈肯的协同学思想,如果把每次挖掘的结果看成一个个微观层次的子系统,那么整个动态挖掘进程就是在已有的结果基础上再进一步地给予宏观上的研究,形成整体地对知识或者规则的认识和评价,有效地防止上一次挖掘的结果的不确定性所带来的影响。动态挖掘进程的这种动态性特点使得利用生物系统的免疫进化解决其中问题成为可能。

3 知识发现中的免疫进化机制内涵

3.1 生物免疫进化过程

在生物免疫系统中,一切免疫反应均源于抗原的入侵。其机理是通过外部抗原的入侵产生免疫应答,引起免疫细胞对抗原的识别,将抗原信息传递给免疫活性细胞;再通过抗体与抗原的匹配(以亲和度度量),对抗体进行选择、变异,促使相应抗体的增殖、分化;之后,引起炎症反应,浆细胞合成并分泌抗体,产生体液免疫。生物免疫系统还可以产生免疫记忆,当再次遇到相同抗原分子时能作出迅速反应,给出更快更强的应答。

生物免疫系统的上述特点成为人们解决工程实际问题的灵感来源。基于生物医学研究对自然免疫系统的不同机理的认识,人们设计出许多利用免疫进化机制处理各种问题的方法。De Castro 和 Von Zoben 开发了基于免疫网络亚动力学的数据分析和数据简化算法[8],并检验了免疫系统内克隆选择机制的作用[9]。该方法继续扩展和应用到数据聚类,产生了 aiNet 算法。实验表明,把进化人工免疫网络与传统统计方法结合,能够有效实现从数据集中提取有用的信息聚类[10]。此外,Timmis 提出了一种资源有限人工分类器[11],并很好地改进了其效率[12]。J. Twycross 提出了基于免疫阴性选择原理的 Web 文本分类算法[13]。

尽管人工免疫系统中的各种模型采用了自然免疫系统的不同机理,但最主要的还是抗原与抗体的亲和作用。对抗体与抗原的亲和作用进行建模几乎在所有的模型和技术中采用。

3.2 动态挖掘进程与生物免疫进化过程的相似性

综合分析知识发现的动态挖掘进程,发现它与生物免疫进化过程有很多相似之处。在此,赋予生物免疫系统的基本概念在动态挖掘进程中新的含义。

1)“抗原”对应动态挖掘中新增的数据；

2)“抗体”对应挖掘得到的知识；

3)“记忆库”对应知识库；

4)“免疫细胞的产生”意味着随机产生或从记忆库中提取知识或规则构成初始群体；

5)“抗原与抗体的相遇”意味着新数据到来后知识的演化；

6)“抗体识别抗原”意味着以知识或者规则的相关值作为个体浓度的适应度，进行抗体适应性的评价，其实，这就是抗体对抗原的结合强度；

7)“抗体的繁殖”意味着基于上一步的计算结果对抗体群体进行选择、变异操作得到新群体；

8)“免疫记忆”意味着把与抗原能够很好匹配的抗体作为规则保存到数据库，以备知识展示或者下一次数据到来前生成初始种群。

图 1 说明了生物免疫进化过程与知识发现系统中动态挖掘进程之间的相似性。

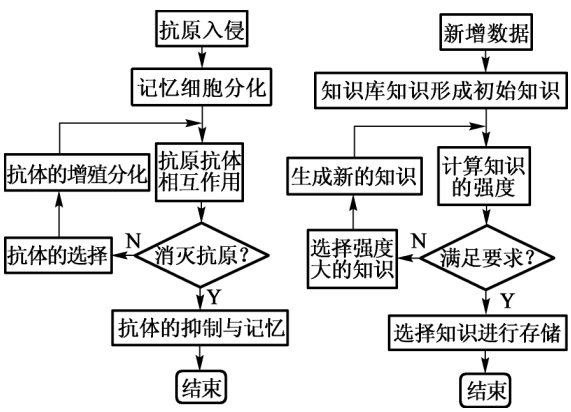

图 1　生物免疫进化过程与动态挖掘进程的相似性比较

Fig. 1　A comparison between the evolutionary immune process and dynamic mining process

3.3　知识发现中的免疫进化机制内涵

综观知识发现的动态挖掘进程与生物免疫进化过程在过程原理上的相似性，并根据实际处理问题的需要，在基于内在认知机理知识发现理论 KDTICM 中的双库协同机制、双基融合机制、信息扩张机制的基础上[14]，提出知识发现系统的免疫进化机制，其基本特征的非形式化描述如下：

1)把新增加的数据作为抗原，已有的知识作为抗体，通过抗体对抗原的识别，依据其结合强度的大小，实现抗体的增殖、分化、变异，通过记录知识的持续数，表征知识的衰减和保持；

2)结合利用双库协同机制中的启发式协调器[14]，把常识、用户以及专家知识作为疫苗，实现对抗体的接种，形成定向挖掘，提高抗体的适应性以及获取新的知识的能力；

3)对获得的新规则不急于作为记忆保存，而是先评价，期望去除矛盾知识，但是重复或冗余的知识要记忆，在形成新的应答抽取初始抗体群时，重复的知识被抽取的可能性更大，宜于实现知识库(记忆库)的实时维护和奉行。

针对动态挖掘进程，展开基于免疫进化机制的知识发现系统的研究极具意义。它通过识别新数据，利用历史挖掘的知识及其参数，实现快速有效的挖掘。利用免疫记忆的原理，对以前经常挖掘出的知识，在一次挖掘中不因为其不能满足参数阈值要求就随意抛弃，而是加以保护(但持续数降低)，这避免了由于数据的随机分布而带来对挖掘结果的过大影响。

4　应用

为了验证知识发现系统中免疫进化机制在解决实际问题中应用的有效性，以时序模式挖掘为例，进行了相关的实验。

将时序数据挖掘看成是一个动态挖掘进程，即在静态时序挖掘的基础上，动态地考察、综合每次挖掘的结果，给出知识的评价；从知识演化的角度，综合地评价知识的类型，给出对知识的取舍。针对随时间变化的时序数据库，在每次挖掘的时候，选用具体的增量式挖掘算法，求得一次挖掘的结果；并利用每次挖掘结果进行动态分析，剔除不确定因素对数据造成的影响，从而改进挖掘结果，指导每次挖掘。

4.1　几个定义

1)规则持续数。

定义 3　对一个规则 R，取 $0<\lambda<1$，称

$$\alpha = q_{i,n} + \lambda(q_{i,n-1} + \lambda(q_{i,n-2} + \cdots + \lambda(q_{i,2} + \lambda(q_{i,1})))) \quad (1)$$

为规则 R 相对第 i 个参数的持续数。

规则持续数反映了规则 R 相对第 i 个参数来说能够持续的强度。实际应用中可以根据需要设定λ的值。由于 $0<\lambda<1$，所以随着挖掘次数的增加，$\lambda^n \to 0$，说明以前挖掘的结果对现在的影响逐渐减弱。

定义 4　在动态挖掘进程中，如果当前是第 n 次挖掘，那么前 $n-1$ 次挖掘得到的规则持续数定义为

$$\beta_i = \alpha_i / ((1-\lambda^n)/(1-\lambda)) \quad (2)$$

β_i 反映了到目前为止，规则应有的持续程度。

在考虑知识的传承和淘汰问题时,可以事先设定规则持续度阈值,然后依据持续度阈值,决定规则的持续还是淘汰。

2)知识的编码。免疫进化机制要求对数据及知识进行必要的编码。对抗原抗体采用实数编码的方式。

设事务数据库有 s 个属性,如果记其第 i 个属性为 X_i,对应的数值域是 D_i,其中 $i = 1, 2, \cdots, s$,则数据库中一条记录就是一个 s 维向量,其每一维上的分量取值对应于相应属性的数值域。它的一个正则划分数值子域为 $D_{i,1}, D_{i,2}, \cdots, D_{i,t}$,定义一个映射为 $D_{i,j} \rightarrow j, j = 1, 2, \cdots, t_i$。

由此映射就可以对一条记录进行如下的编码:如果这条记录的第 i 个属性取 $D_{i,j}$,那么其编码的第 i 位就取得 j 值;如果这条记录的第 i 个属性值取空值,则该位置取 0。这样,一条记录其编码的第 i 位就是 0, 1, …, t_i 中的某一个确定的值了,即一条记录的编码第 i 位的值域是 $\{0, 1, \cdots, t_i\}$,记为 V_i,即 $V_i = \{0, 1, \cdots, t_i\}$。

定义 5　称事务数据库 X 对应的所有属性,其对应的编码值域 V_i 的笛卡尔积

$$V = V_1 \times V_2 \times \cdots \times V_s \tag{3}$$

为抗原的形态空间。

对于抗体(频繁项目集或者候选项目集)的编码可以采用同样的编码原则,因为频繁项目集或者候选项目集的长度一般都小于记录的属性个数,这样只要把空缺位置填 0,把项目集对应属性位的值对应于相应的编码就可以得到抗体的编码。

3)抗体与抗原的结合强度和抗体对抗原的刺激水平。计算抗体与抗原的结合强度的一般方法包括:海明空间的海明距离、Eucilidean 形态空间的 Eucilidean 距离和 Manhattan 形态空间的 Manhattan 距离等。根据知识发现系统的动态挖掘进程的特点,结合上述的抗原、抗体的编码形式,定义抗体、抗原的结合强度如下:

定义 6　如果抗原的编码是 $\mathrm{Ag} = (\mathrm{ag}_1, \mathrm{ag}_2, \cdots, \mathrm{ag}_L)$,抗体的编码是 $\mathrm{Ab} = (\mathrm{ab}_1, \mathrm{ab}_2, \cdots, \mathrm{ab}_L)$,取

$$\delta = \begin{cases} 1 & \text{if } \mathrm{ag}_i = \mathrm{ab}_i \\ 0 & \text{else} \end{cases} \qquad \mu = \begin{cases} 1 & \text{if } \mathrm{ab}_i \neq 0 \\ 0 & \text{else} \end{cases}$$

其中 $i = 1, 2, \cdots, L$,那么,抗体 Ab 与抗原 Ag 的结合强度为

$$D = \sum_{i=1}^{L} \delta_i \Big/ \sum_{i=1}^{L} \mu_i \tag{4}$$

当 $D = 1$ 时,抗体与抗原就是完全匹配的。

定义 7　把对一个抗体与所有抗原的结合强度相加,得到的值称为所有抗原对抗体的刺激水平。刺激水平可以决定对抗体的取舍。

定义 8　假设有两个时间序列 $\boldsymbol{A}_1 = \{x_{11}, x_{12}, \cdots, x_{1n}\}$ 和 $\boldsymbol{A}_2 = \{x_{21}, x_{22}, \cdots, x_{2n}\}$,其模式分别为 $\{a_{11}, a_{12}, \cdots, a_{1n}\}$ 和 $\{a_{21}, a_{22}, \cdots, a_{2n-1}\}$,如果满足 $d(\boldsymbol{A}_1, \boldsymbol{A}_2) \leqslant \mathrm{pc}$,则称时间序列 $\boldsymbol{A}_1$ 与 $\boldsymbol{A}_2$ 是在给定误差 pc 允许下的相似时序模式。

相似关系一般应该满足传递性,但上述定义的相似模式不满足传递性,为相似关系的处理带来了不便。另外,考虑到下一步编码的方便,给出了相似模式的改进定义。

定义 9　把 -90°到 90°之间 n 等分。如果一条线段对应的编码等于 i,则它相似于角度 $-90° + 180°/2n + i \times 180°/n$ 对应的模式。如果 2 条线段对应的编码相等,则它们相似。

定义 9 满足相似关系的传递性,即相似于同一模式的 2 个线段,本身也相似,从而具有传递性,为处理相似关系带来方便。该定义也为编码带来方便。编码采用十进制,一个线段模式对应的编码就是它最接近的中心角的编码值,一个序列的模式就可以定义为它的每一个线段对应的编码序列。然后,把要找的候选模式作为抗体,把时间序列对应的窗口序列集中的序列作为抗原,并根据式(4)将抗体与抗原之间的结合强度进一步定义如下:

定义 10　如果抗原的编码是 $\mathrm{Ag} = (\mathrm{ag}_1, \mathrm{ag}_2, \cdots, \mathrm{ag}_L)$,抗体的编码是 $\mathrm{Ab} = (\mathrm{ab}_1, \mathrm{ab}_2, \cdots, \mathrm{ab}_L)$,取

$$\delta = \begin{cases} 1 & \text{if } \mathrm{ag}_i = \mathrm{ab}_i \\ 0 & \text{else,} \end{cases}$$

这里 $i = 1, 2, \cdots, L$,那么,抗体 Ab 与抗原 Ag 的结合强度为

$$W = \sum_{i=1}^{L} \delta_i \Big/ L \tag{5}$$

当 $W = 1$ 时,抗体与抗原完全匹配,在计算结合强度时,可以作为该序列的支持度。

4.2　算法

基于免疫进化机制的时序模式发掘算法:

输入 n(把 -90°到 90°等分为 n 份);窗口 w;时间序列 $\boldsymbol{A} = \{x_1, x_2, \cdots\cdots, x_n\}$;

输出时序模式。

Step1　先由时间序列和窗口 w 值,得到对应的

序列集 $\boldsymbol{W}(s)=\{s_1,s_2,\cdots,s_{n-w+1}\}$；

Step2　计算序列集中每一个序列的模式作为抗原 Ag，得到编码；

Step3　随机取若干个抗原作为抗体，产生初始抗体种群 Abs，放入记忆矩阵 $\boldsymbol{M}$；

Step4　亲和度计算，根据抗原对抗体的刺激水平，计算当前抗原 Ag 和 $\boldsymbol{M}$ 中每个抗体 Ab 的结合强度，统计支持度；

Step5　克隆选择，选择具有高结合强度的抗体进行克隆，规模正比于结合强度；

Step6　抗体成熟度，抗体进行变异，规模反比于结合强度，接种疫苗；

Step7　重新选择：计算每个 Ab 和当前抗原 Ag 的结合强度，重新选择具有较高支持度的抗体，将具有较低支持度的抗体作为当前抗原；

Step8　重复执行 Step5 ~ Step7，直到满足规定的执行次数；

Step9　if 第一次计算 then 转 Step11；

Step10　进行抗体抑制，对新出现的支持度超过最小支持度的抗体，到原来数据中进行检查；

Step11　分类记忆结合强度高的抗体，以及抗体持续数；

Step12　展示支持度大于最小支持度的模式；

Step13　监视直到新数据的到来；

Step14　将新数据结合原来的数据，生成新的窗口序列集合 $\boldsymbol{W}(s)=\{s_1,s_2,\cdots,s_{n-w+1}\}$（抗原）；

*Step*15　由记忆抗体集产生新的初始抗体种群 Abs，放入矩阵 $\boldsymbol{M}$，转入 Step4；

Step16　重复执行 Step4 ~ Step15，直到满足结束条件；

Step17　结束。

4.3　实验分析

取某股票 8 个月中每天的平均价作为实验数据（见图 2）。取 $w=4$，观察连续 4 天股价的变化形态。取 $n=9$，允许最大偏差为 20°，支持度取 0.15。

先取 3 个月数据运行程序，分析得到 2 个模式（见图 3）。模式显示出股价以上涨和调整为主。然后，增加 2 个月的数据，系统很快显示这 2 个模式仍为频繁模式，说明股价仍以上扬为主。又增加 3 个月数据后，系统显示这 2 个模式仍是频繁模式，但是同时出现另外一个下降模式（见图 4）。考虑到下降模式出现较晚，并且其持续度比上升模式持续度较低，所以总的来说可以认为股价呈现上扬调整的态势。

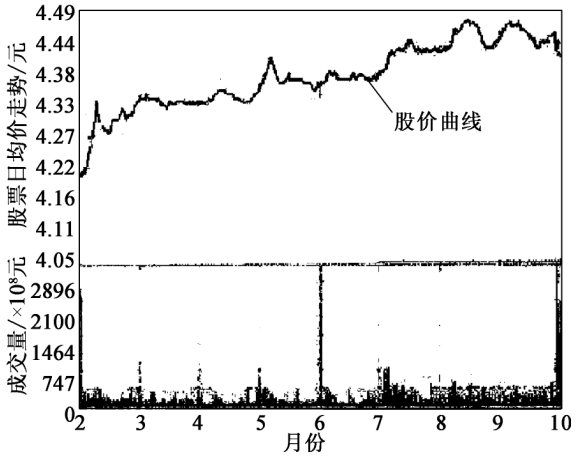

图 2　某股票 8 个月每天交易的平均价格走势

Fig. 2　The average price trend for a certain stock

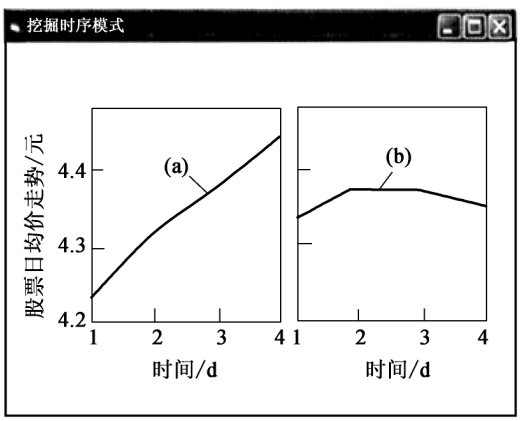

图 3　第一次、第二次显示模式

Fig. 3　Patterns shown at the first and second runtime

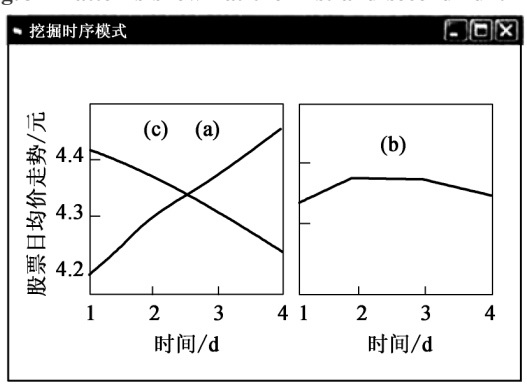

图 4　第三次显示模式

Fig. 4　Patterns shown at the third runtime

由此可见，基于数据增量的动态挖掘进程利用已有模式及其变异进行发现，不仅可以较快发现新的模式，而且它在对知识的评价运用上，不是依据前一次挖掘的结果，而是根据每一次挖掘结果，综合地历史地对挖掘的规则进行评价。这种历史地综合地跟踪性研究能够避免一次挖掘中数据随机因素对挖

掘结果造成的干扰,说明利用免疫进化机制解决动态挖掘问题的方法是有效的。

5 结语

在阐明数据挖掘中动态挖掘进程与生物免疫进化机制之间相似性的基础上,提出知识发现系统中的免疫进化机制内涵,并通过时序模式挖掘验证了新机制的有效性和先进性。由于生物免疫进化机制是复杂的,所以一般应用系统都是从某一特定角度出发利用它来解决实际工程问题。今后,一方面要继续跟踪生物免疫进化方面的新研究进展,另一方面要扩充和完善现有的知识,发现系统中免疫进化机制的理论基础,将其应用到更多类型的数据挖掘进程中。

参考文献

[1] Cheung D W, Han J. A fast algorithm for mining association rules [A]. Proceedings of the 4th International Conference on Parallel and Distributed Information System [C], Miami Beach, Florida, 1996:73 - 84

[2] 何炎祥, 张 戈,石 莉. 关联规则的维护[J]. 计算机工程与应用, 2002, 25(10): 203 - 205

[3] 冯玉才, 冯剑琳. 关联规则的增量式更新算法[J]. 软件学报, 1998, 9(4): 301 - 306

[4] 宋余庆, 朱玉全, 孙志挥. 基于 FP - tree 的最大频繁项目集挖掘及更新算法[J]. 软件学报, 2003, 14(9): 1586 - 1592

[5] 杨 明, 孙志挥. 一种基于前缀广义表的关联规则增量式更新算法[J]. 计算机学报, 2003, 26(10): 1318 - 1325

[6] 吉根林, 杨 明, 宋余庆. 最大频繁项目集的快速更新[J]. 计算机学报, 2005, 28(1): 128 - 135

[7] Agrawal R, Srikant. Fast algorithms for mining association rules [A]. Proceedings of the 20th International Conference on Very Large Databases [C]. Santiago, Chile, 1994:487 - 499

[8] Castro De, Timmis J. Artificial immune systems: a novel approach to pattern recognition [A]. Alonso L, Corchado J, Fyfe C eds. Artificial Neural Networks in Pattern Recognition [C]. University of Paisley, U K, 2002:39 - 50

[9] Castro De, Von Zuben F. The clonal selection algorithm with engineering applications [A]. Proceedings of Genetic and Evolutionary Computation Conference [C], Berlin Heidelberg: Springer - Verlag, 2000:121 - 132

[10] Castro De, Von Zuben F. An evolutionary immune network for data clustering [A]. Proceedings of the IEEE Computer Society Press SBRN001 [C]. 2000:84 - 89

[11] Timmis J, Neal M. A resource limited artificial immune system for data analysis [J]. Knowledge Based on Systems, 2001, 14(34): 121 - 130

[12] Watkins A, T immis J. Artificial Immune Recognition System (AIRS): Revisions and Refinements [M]. Berlin Heidelberg: Springer - Verlag, 2003

[13] Twycross J. An Immune System Approach to Document Classification [D]. HP Laboratories Bristol, 2002:189 - 199

[14] 杨炳儒. 知识发现与知识工程[M]. 北京: 冶金工业出版社, 2000

Evolutionary Immune Mechanism and Its Application on Temporal Sequential Pattern Mining

Yang Bingru[1], Qin Yiqing[1,2], Song Zefeng[1]

(*1. College of Information Engineering, Beijing University of Science and Technology, Beijing 100083, China;*
2. Department of Computer and Automation, Beijing Institute of Machinery, Beijing 100085, China)

[**Abstract**] A new approach to solve the problem of dynamic data mining is presented. Firstly a new concept of dynamic mining process is proposed. Next the evolutionary immune mechanism in KDD is illustrated, based on a comparison between the dynamic mining process and biological immune process. Additionally how to apply the approach to temporal sequential pattern mining and evaluate the experimental results are described. Finally the work and present proposals for future work are concluded.

[**Key words**] dynamic data mining; immune algorithm; dynamic mining process; evolutionary immune mechanism; temporal sequential pattern mining