使用二次连接神经网络的基于 ART 的分层聚类算法*)

顾 明[1,2]

(深圳职业技术学院软件工程系 深圳 518055)[1]

(哈尔滨工业大学软件工程有限公司博士后工作站 哈尔滨 150001)[2]

摘 要 本文描述了二次连接神经网络的结构和特性,给出了该网络的非监督学习规则。使用二次连接的神经网络,描述了基于 ART 的层次聚类算法,并对算法的时间和空间复杂性进行了分析。为了检验算法的有效性,采用了一个人工的二维数据集合,并比较了本文提出的算法和具有代表性的 K-means 算法的执行效果。

关键词 神经网络,聚类分析,非监督学习,算法

ART-based Hierarchical Clustering Algorithm Using Quadratic Junction Neural Networks

GU Ming[1,2]

(Dept. of Software Engineering, Shenzhen Polytechnic, Shenzhen 518055)[1]

(Postdoctor Workstation of Software Engineering Limited Corporation, Haerbin Industry University, Haerbin 150001)[2]

Abstract In this paper, structure and properties of neural networks with quadratic junction are presented. Unsupervised learning rules about the neural networks are given. Using this kind of neural networks, an ART-based hierarchical clustering algorithm is suggested. The time and space complexity of the algorithm is discussed. A 2-D artificial data set is used to illustrate and compare the effectiveness of the proposed algorithm and K-means algorithm.

Keywords Neural network, Cluster analysis, Unsupervised learning, Algorithm

1 引言

在聚类算法中,有两个主要的问题需要解决:一个是优化簇数的确定,这被称为簇有效性问题;另一个是相似性测量,这意味着根据什么原则来安排输入模式到相应的簇。解决第一个问题的传统方法是增加或合并存在的簇数,每次增加或合并后都计算簇有效性,直到获得优化的簇数[1~4]。因为大部分有效性测量方法都涉及到簇形状有一定的几何结构的问题,所以这些方法在估算正确的簇数方面都有一定的局限性。相似性测量面临的问题是簇的一般形式化定义。大部分传统的簇算法假定,有相似位置或连续密度的模式属于一个簇。为了从数学上识别数据集中的簇,首先需要定义相似性(亲近性)的测量。相似性是建立一些规则来指定某些模式到特定的簇中心域,这意味着不同的相似性测量将导致不同的聚类结果。

已经有许多方法来解决前面提到的两个问题[5~9],然而没有一个方法可以对任何应用都适用。簇的寻找是面向应用的,而且簇算法或多或少依赖某些参数植。一种解决方案是提供一个补充的方法,使数据分析员根据情况确定合适的簇数。

人工神经网络可以应用在数据聚类方面,再结合分层可以为聚类提供一个有效的解决方案。

本文结合分层聚类的优点和 adaptive resonance theory (ART),神经网络不需事先说明簇数的优点,来增加聚类结果的优化程度。其中二次连接神经元的概念来自文[10],这类神经元组成的网络可以聚类数据成为超椭圆形。该文没有提供簇的初始化条件,文本提出在训练过程中动态增加神经元数量,并且依次在每个输入模式中初始化神经元权重 $b'_j s$,可以防止有些初始化不好的神经元得不到学习和竞争胜利的机会。对文[11]的有些思想进行了优化,产生的第一层系统树状图与其余的层分离,这样可以极大地改善算法的时间和空间复杂度。文[12]用 inverse squared 距离来测量相似性,而我们的算法则使用二次神经元的输出函数来测量,这与 Mahalanobis 距离有着相似的影响。另外,存储在簇-检测-标志网络中的原型在文[12]中实际上是某些前面的输入模式,而我们的算法以可修改的方式存储这些原型,可修改性通过更新获胜二次神经元的权重的方式完成。

本文组织如下:第 2 节描述了使用二次连接的神经网络;第 3 节具体说明了分层的聚类算法;第 4 节呈现了一个二维人工数据集的实验模拟结果以及与 K-means 算法的比较;最后是总结和进一步的研究方向。

2 使用二次连接的神经网络

2.1 结构和特性

在神经网络的应用中,最经常遇到的神经元类型连接是线性和加上 S 类的激活函数。研究表明,具有 S 输出函数的三层感知器是通用的模拟器,这意味着可以训练模拟器模拟任何输入和输出之间的映像。模拟器的精确性取决于隐藏层的神经元个数。然而,每层神经元的个数和参数 η 和 α 的最佳值的关系一直没有很好地解决,处理得不好,可能会导致神经网络花费很长的时间得到解,或者甚至没有解。

可以考虑以另一种方式模拟神经元之间的连接[13,14]。事实上,生物神经元本身是非线性的过程。换句话说,线性连接对于表示发生在神经元之间的关系,并不是一个足够好的

*)深圳市科技计划项目基金资助(05KJCD020)。顾 明 副教授,博士,主要研究方向:软件工程、网络安全、人工神经网络。

选择。二次神经元的优点是能够捕捉高级的相互关系，虽然具有足够多的隐藏结点的多层线性神经元也能够完成非线性的映像，但与二次神经元相比，多层线性神经元使用较多的系数，并且具有较低的学习速率。

二次连接神经元的描述公式和结构如下：

$$y_{jk}(x)=\sum_{i=1}^{n} w_{ji_k} x_i \tag{1}$$

$$net_j(x)=\sum_{k=1}^{n}(y_{jk}(x)-b_{jk})^2 \tag{2}$$

$$Out_j(x)=e^{-s_j^2 net_j(x)} \tag{3}$$

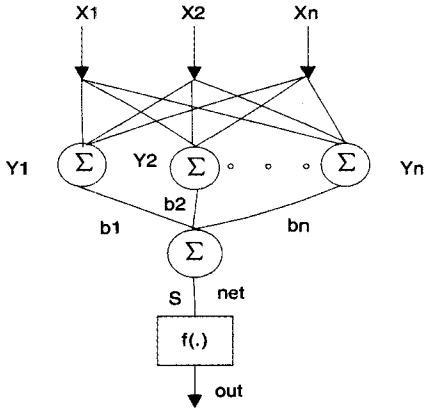

图 1　二次神经元的结构

式中 b_{jk}，s_j 和 w_{jk_i} 是可调整的权重，$x=(x_1,x_2,\cdots,x_n)^T$ 是一个输入模式，$y_j=(y_{j1},y_{j2},\cdots,y_{jn})^T$ 是输入模式 x 的转置，$Out_j(x)$ 是神经元 j 的输出函数。

二次连接神经元的结构能够分成两部分：一部分由公式(1)给出的线性连接组成，完成输入数据的线性转换；另一部分由公式(2)给出，产生相应的超椭圆型。这种类型的二次连接能够完成在大小和位置上都可以变化的超椭圆的判别。

2.2　学习规则

$$b_{jk}(n+1)=b_{jk}(n)+\eta_b(2s_j^2 Out_j(x)(y_{jk}(x)-b_{jk})) \tag{4}$$

$$s_j(n+1)=s_j(n)+\eta_s(-2s_j^2 Out_j(x)net_j(x)) \tag{5}$$

$$w_{jk_i}(n+1)=w_{jk_i}(n)+\eta_w(-2s_j^2 Out_j(x)(y_{jk}(x)-b_{jk}x_i)) \tag{6}$$

式中 η_b，η_s 和 η_w 是正的学习速率，通过使用更新权重的相同输入模式的重复，规则增加了第 j 个神经元获胜的机会。

对于训练网络，初始化权重向量是重要的。初始化不当的神经元可能没有获胜的机会，它们也失去了学习的机会，这将会导致不能发现合适的簇。可以用以下三种方式解决初始化权重问题：

(1)初始化权重向量 b_{jk} 为输入模式本身的采样值；

(2)不仅仅更新获胜神经元的权重，也更新未获胜神经元的权重；

(3)增加一种控制机制，减少经常获胜神经元再次获胜的机会。

以下的算法采用(1)的方式。

3　基于 ART 的分层聚类算法的基本思想

Adaptive Resonance Theory (ART)神经网络接受输入模式，产生对输入模式的分类代码。通过增加一个新的神经元到输入层，ART 神经网络可以创建一个新类，因此不需要预先说明簇数，ART 神经网络便能够聚类数据[15,16]。为了处理簇数的问题，ART 神经网络采用警戒参数的方式，来决定什么时候产生新簇。尽管 ART 神经网络不能指定簇数，但在某种程度上，警戒参数可以隐含地预先说明簇数。即：警戒参数越大，簇数越小。关键问题是，簇数依赖于警戒参数的值，由此而产生的聚类结果的满意度如何？需要提供一个方法来确定不同的警戒参数值得到的簇数，哪个是最满意的。

分层聚类方法能够解决以上问题，能够不断合并簇的分层聚类，被称为收敛的。分层聚类结果由系统树图表示。该系统树图由一些结点组成，每个结点表示一个簇，连接结点的线表示簇的合并，每层表示数据集的聚类结果。通过检测系统树图的结构，可以在基于 ART 的聚类算法产生的结果中选择一个满意的聚类结果。

我们建议的算法分成两个阶段：第一阶段是建立一层具有二次连接神经元的神经网络，这种神经元能够把数据分类成超椭圆型，二次神经元的输出显示了当前输入模式和已在超椭圆型内的神经元组成的原型之间的相似性。二次神经元的输出可以通过公式(1)～(3)来计算。我们用阈值作为相似性度量的函数，阈值的作用相当于警戒值在 ART 中的作用。当产生的二次神经元的输出小于预先指定的阈值时，就产生一个新的二次神经元；如果大于阈值，获胜神经元的权重被更新，以便使得相应的椭圆型适合于输入模式。重复这个过程，直到处理完所有的输入数据(模式)。

第二个阶段是减小阈值，对某个数据模式依次检测，是否有某些神经元的输出还大于已经减小的阈值，并形成“同类集”。在“同类集”中的二次神经元将属于相同的簇。不断减小阈值，直到生成系统树图的结构为止。

最后，通过检测生成的系统树图，来确定合适的簇数。对于不同的应用，通过检测系统树图，可以得到不同的聚类结果。

3.1　建立第一层神经网络的算法描述

步骤 1.1：选择一个初始的警戒参数值 θ 和相应的学习参数值 η_b，η_s 及 η_w。

步骤 1.2：如果当前的输入模式是数据集中的第一个模式，用公式(7)，(8)和(9)来初始化二次神经元的权重。

$$b_j(0)=x \tag{7}$$

$$w_{jk_i}(0)=\begin{cases}1 & \text{if } k=i\\ 0 & \text{其它}\end{cases} \tag{8}$$

$$s_j=1 \tag{9}$$

步骤 1.3：使用基于 ART 的最大值标准，在已经产生的神经网络的神经元中，依据以下公式(10)寻找获胜神经元 $j^*(x)$：

$$j^*(x)=\underbrace{\operatorname{argmax}}_{j=1,2,\cdots,J} Out_j(x) \tag{10}$$

式中 J 是当前产生的神经网络中神经元的个数。

步骤 1.4：如果获胜神经元 $j^*(x)$ 的输出大于阈值，即 $Out_{j^*}(x)\geqslant\theta$，则使用公式(4)，(5)和(6)调整获胜神经元 $j^*(x)$ 的权重。否则，产生一个新的神经元，并用以下公式(11)和(12)初始化权重：

$$b_{j+1}(0)=x \tag{11}$$

$$w_{(j+1)k_i}(0)=\begin{cases}1 & \text{if } k=i\\ 0 & \text{其它}\end{cases} \tag{12}$$

步骤 1.5：处理下一个输入模式，重复步骤 1.3 和 1.4，直到处理完所有的输入模式为止。

3.2　分层聚类第一层的创建算法描述

定义 1　同类集：矩阵 E_{k+1} 的行列为 $N\times J$，矩阵中的元

素为 e_{ij} ($1\leqslant i\leqslant N, 1\leqslant j\leqslant J$)，$e_{ij}$ 的取值为 1 或 0。对于两个二次神经元 p 和 m，如果对某个输入模式 x，存在 $e_{ip}=e_{im}=1$，则这两个二次神经元属于相同的同类集。

同类集的直观解释是：属于相同同类集的神经元应该属于相同的簇。

定义 2 S_l^{k+1} 表示第 l 个同类集，N_c^{k+1} 表示在第 $(k+1)$ 次循环中形成的同类集的数量。

例如，在图 2 中，在数据集中有 5 个输入模式，已经产生的神经元数量为 6，形成的矩阵为 E_1。根据以上的定义，可以发现有两个同类集 $S_1^1=\{1,3,5\}$ 和 $S_2^1=\{2,4,6\}$。

$$E_1=\begin{array}{c} \\ 1\\2\\3\\4\\5\end{array}\begin{array}{c}\begin{array}{cccccc}1&2&3&4&5&6\end{array}\\ \begin{pmatrix}0&1&0&1&0&0\\0&1&0&0&0&1\\1&0&1&0&0&0\\0&0&1&0&1&0\\0&1&0&1&0&1\end{pmatrix}\end{array}$$

图 2　同类集的例子

步骤 2.1：设置循环次数 k 为 0，确定参数 α 的值，使 α 的范围是(0,1)之间。初始化一个 $N_k\times J_k$ 矩阵 E_{k+1} 的所有元素为 0，其中 N_k 和 J_k 分别表示输入模式的数量和在第一阶段产生的神经元的数量。

步骤 2.2：依据以下公式(13)计算矩阵中 E_{k+1} 的元素：

$$e_{ij}=\begin{cases}1 & \text{if } Out_j(x_i)\geqslant\alpha^{k+1}\theta\\ 0 & \text{其它}\end{cases}\qquad(13)$$

步骤 2.3：通过检测矩阵 E_{k+1}，找出同类集，假设找到的同类集的数量是 J_{k+1}。

步骤 2.4：对于每一个同类集，找到一个输入模式，该输入模式使得该同类集中的神经元具有最小的输出。用这个输入模式代表该同类集中的所有输入模式。此时，同类集的数量 J_{k+1} 等于输入模式的数量。

步骤 2.5：使 k 值增加 1，初始化矩阵 $J_{k+1}\times J_{k+1}$ 的所有元素为 0。

步骤 2.6：重复步骤 2.2～2.5，直到循环数达到预先定义的值，或者同类集 N_c^{k+1} 的数量成为 1。

3.3　算法的复杂性分析

在第一阶段，算法的时间复杂度是 $O(n\cdot J)$，其中 n 代表数据集中输入模式的维度，J 代表产生的神经元数量；对步骤 1.2 和 1.4，算法的空间复杂度是 $O(n)$；对步骤 1.3，算法的空间复杂度是 $O(J)$。

在第二阶段，算法的时间和空间复杂度都为 $O(N_k,J_k)$，其中 N_k 和 J_k 分别表示输入模式的数量和同类集的数量。它们随着 k 的增加而减小，在每次循环中取不同的值。

4　模拟实验的结果

我们用一个二维的模拟数据集来检测算法的有效性，因为二维数据集容易直观地看到聚类后的效果，但并不意味着算法仅适用于二维数据。对三维及其以上的数据，算法也同样适用。

图 2 是由 186 个采样数据组成的任意形状的输入模式，我们用 VisualC^{++} 6.0 设计和实现了该算法来演示结果。参数的取值 $\theta=0.7$，$\alpha=0.8$，$\eta_p=\eta_s=\eta_w=0.006$，软件的用户界面如图 3 所示，算法生成的系统树状图如图 4 所示。在图 4 的第一层，显示在算法的第一阶段，产生了 18 个二次神经元(簇)；在算法的第二阶段，18 个二次神经元被合并成 3 个簇，然后 3 个簇被合并成 2 个簇，最后 2 个簇被合并成一个簇。

通过观察图 4 的系统树状图，我们可以做出这样的决定：该数据集可以有 2 个或 3 个簇。在表 1 中，比较了本文的算法和 K-means 算法的聚类结果。

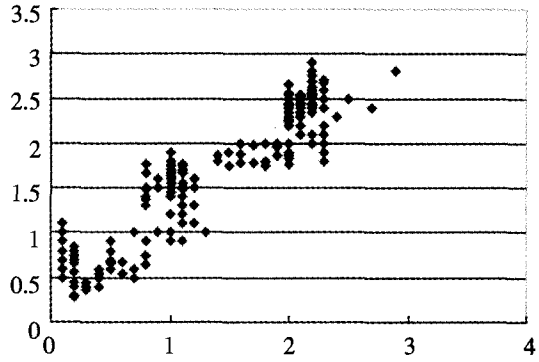

图 2　采样数据为 186 的二维输入模式

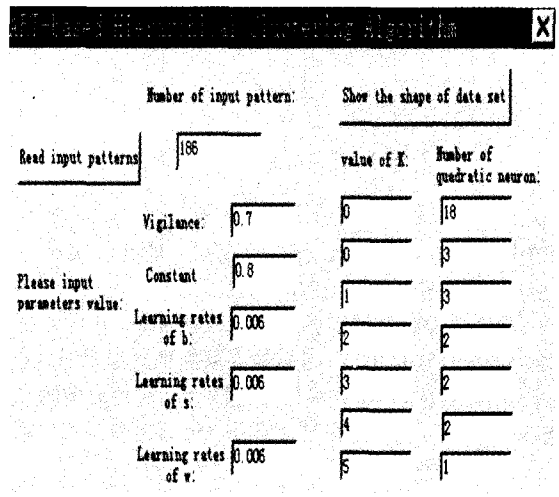

图 3　实现算法软件的用户运行界面

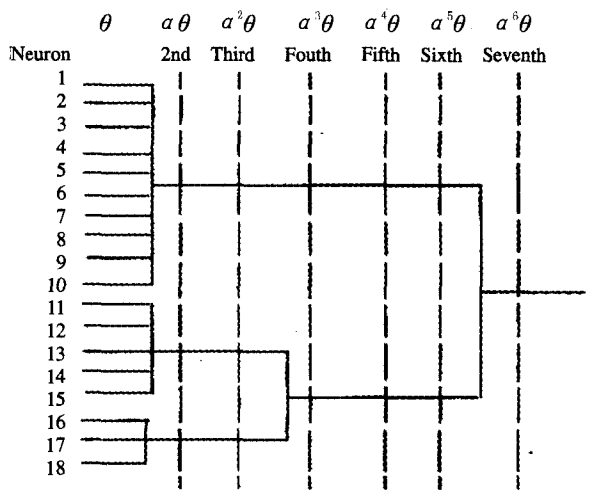

图 4　算法生成的系统树状图结构

结论和进一步的工作　ART 神经网络具有不需要预先指定簇数量和能够提供最大簇数量的优点，但是这些性能依赖于警戒参数的值。层次聚类技术可以使数据分步组成簇，可以为数据分析员提供直观的数据图示结果。层次聚类技术

的局限性是生成的系统树状图是很多种模式组合的结果，这使分析员很难决定最终应该选择哪一种。我们提出的算法结合这两种方法的优点，既克服了 ART 神经网络不同的参数可导致不同结果的局限性，又用系统树状图来辅助可视化地选择簇数量。

表 1　本文算法和 K-means 算法的比较

算法 / 结果	K-means 簇数=3	K-means 簇数=2	该算法 簇数=3	该算法 簇数=2
误聚类的模式数/输入模式数	14/186	5/186	2/186	0/186

聚类算法的重要的问题是结合应用领域来聚类数据。因此，我们进一步的工作包括两个方面：一个方面是结合应用领域，使我们的算法应用于大的数据集并聚类任意不规则的数据分布形态，因为使用二次连接的神经网络能够处理任意的数据分布形态，所以聚类任意不规则的数据形态也是可实现的。另一方面是结合不同的应用领域，为算法的参数选择合适的值。

参 考 文 献

1 Krishnapuram R, Kim J. A note on the Gustafson-Kessel and adaptive fuzzy clustering algorithms. IEEE Trans on Fuzzy Systems, 1999, 7(4): 453～461

2 Dave R N. Fuzzy shell-clustering and application to circle detection in digital images. Intern Journal of Genera Systems, 1990, 16: 343～355

3 Dave R N. Use of the adaptive fuzzy clustering algorithm to detect lines in digital images. Intell Robots Comput Vision VIII, 1989, 1192: 600～611

4 Gath I, Geva A B. Unsupervised optima Fuzzy Clustering. IEEE Trans on Pattern Analysis and Machine intelligence, 1989, 11: 773～781

5 Eltoft T, de Figueiredo R J P. A new neural networks for cluster-detection-and-labeling. IEEE Trans. on Neural Networks, 1998, 9(5): 021～1035

6 Geva A B. Hierarchical unsupervised fuzzy clustering. IEEE Trans on Fuzzy Systems, 1999, 7(4): 723～733

7 Su M C, Chang H C. A new model of self-organizing neural networks and its application in data projection. IEEE Trans on Neural Networks, 2001, 112(1): 153～158

8 Su M C, Chou C H. A modified version of the k-means algorithm with a distance based on cluster symmetry. IEEE Trans on Pattern Analysis and Machine Intelligence, 2001, 23(6): 674～680

9 Su M C, Chou C H. A competitive learning algorithm using symmetry. IEEE Trans on Fundamentals of Electronics, Communications and Computer Science, 1999, E82-A(4): 680～687

10 Su M C, Liu T K. Application of neural networks using quadratic junctions in cluster analysis. Neurocomputing, 2001, 37: 165～175

11 Su Mu-Chun, Liu Yi-Chun. A hierarchical approach to ART-like clustering algorithm. Neural Networks. In: 2002. IJCNN '02. Proceedings of the 2002 Internationl Joint Conference, vol1, May 2002. 788～793

12 Eltoft T, deFigueiredo R J P. A new neural networks for cluster-detection and-labeling. IEEE Trans on Neural Networks, 1998, 9(5): 1021～1035

13 DeClaris N, Su M C. A novel class of neural networks with quadratic junction. In: IEEE International Conference on Systems, Man, and Cybernetics, 1991. 1557～1562

14 DeClaris N, Su M C. Introduction to the theory and application of neural networks with quadratic junctions. In: IEEE International Conference on Systems, Man, and Cybernetics, 1992. 1320～1325

15 Carpenter G, Grossberg S. Adaptive resonance theory: stable self-organization of neural recognition codes in response to arbitrary lists of input patterns. In: Proc. 8th Annu Conf Congnitive Sci Soc,. 1986. 45～62

16 Carpenter G, Grossberg S, Rosen D B. Fuzzy ART: Fast Stable learning and Categorization of Analog Pattern by an Adaptive Resonance System. Neural Networks, 1991, 4: 759～771

（上接第 127 页）

20 Adibi J, Chalupsky H, Melz E, et al. The KOJAK Group Finder: Connecting the Dots via Integrated Knowledge-Based and Statistical Reasoning. In: AAAI, 2004. 800～807

21 Adibi J. Link Discovery via a Mutual Information Model: From Graphs to Ordered Lists. In: DIMACS Workshop on Applications of Order Theory to Homeland Defense and Computer Security, Rutgers, 2004

22 Upal M A. Performance Evaluation Metrics for Link Discovery Systems. In: Proceedings of the Third International Conference on Intelligent Systems Design & Applications, Springer Verlag, New York, 2003. 273～282

23 Weiss S, Kulikowski C. Computer Systems That Learn Classification and Predication Methods from Statistics, Neural Networks, Machine Learning, and Expert Systems. San Mateo, CA: Morgan Kaufmann, 1991

24 Upal M A, Neufeld E. Comparison of nonhierarchical Unsupervised Classifiers. In: Proceedings of the International Conference on Information, Statistics and Induction in Science, Singapore: World Scientific, 1996

25 IET. Performance Evaluation Specifications for EELD: [Technical Report]. Information Extraction & Transport Inc. 2002 (http://www.iet.com/Projects/EELD)

26 Davision A C, Hinkley D V. Bootstrap Methods and their Application. Boston: Cambridge University, 1997

27 http://infowar.net/tia/www.darpa.mil/iao/EELD.htm, 2005-4-12

28 MckKay S J, Woessner P N, Roule T J. Evidence extraction and link discovery (EELD) seedling project, database schema description. (version 1.0): [Technical Report]. 2862. Veridian Systems Division, 2001

29 http://www.rl.af.mil/tech/programs/eeld/, 2005-4-12

30 Kovalerchuk B, Vityaev E. Correlation of complex evidences and link discovery. In: The Fifth International Conference on Forensic Statistics, Venice, 2002

31 Cook D J, Holder L B. Graph-based data mining. IEEE Intelligent Systems, 2000, 15(2): 32～41

32 Friedman N, Getoor L, Koller D, et al. Learning probabilistic relational models. In: Proceedings of the 16th International Joint Conference on Artificial Intelligence, 1999. 1300～1307

33 Kramer S, Lavrac N, Flach P. Propositionalization approaches to relational data mining. Relational Data Mining. Berlin: Springer Verlag, 2001. 262～291

34 Neville J, Jensen D. Iterative classification in relational data. In: Papers from the AAAI-00 Workshop on Learning Statistical Models from Relational Data, Austin, TX: AAAI Press/ The MIT Press, 2000

35 Zhou Zhi-hua. Three perspectives of data mining. Artificial Intelligence, 2003, 143: 139～146

36 Mack R, Hehenberger M. Txt-based knowledge discovery: search and mining of life-sciences documents. Drug Discovery Today, 2002, 7(11): 89～98

37 Hosking J R M, Pednault E P D, Sudan M. A statistical perspective on data mining. Future Generation Computer Systems, 1997, 13: 117～134

38 Aggarwal C C, AL-Garawi F, Yu P S. Intelligent crawling on the World Wide Web with arbitrary predicates. In: Proc. ACM WWW, 2001

39 Toyoda M, Kitsuregawa M. Creating a Web community chart for navigating related communities. In: Proc. ACM HT, 2001

40 Kovalerchuk B, Vityaev E, Ruiz J F. Consistent and Complete Data and "Expert" Mining in Medicine. In: Medical Data Mining and Knowledge Discovery, Springer, 2001. 238～280