

# 分类方法的新发展:研究综述<sup>\*</sup>)

张丽娟<sup>1</sup> 李舟军<sup>2</sup>

(国防科学技术大学计算机学院 长沙 410073)<sup>1</sup>

(北京航空航天大学计算机科学与工程学院 北京 100083)<sup>2</sup>

**摘要** 分类是数据挖掘的重要任务之一,也是机器学习、模式识别和人工智能等相关领域广泛研究的问题。分类在实际中有广泛的应用,包括医疗诊断、信用评估、选择购物等。近年来,随着相关领域中新技术的不断涌现,分类方法也得到了新发展。本文对这些新发展进行了较详细的归纳,总结了分类方法发展的趋势。

**关键词** 分类,数据挖掘,人工智能,机器学习

## The New Developments of Classification Methods: Overview

ZHANG Li-Juan<sup>1</sup> LI Zhou-Jun<sup>2</sup>

(Computer School, National University of Defence Technology, Changsha 410073)<sup>1</sup>

(School of Computer Science & Engineering, Beihang University, Beijing 100083)<sup>2</sup>

**Abstract** Classification is one of the important tasks in data mining. It has also been extensively investigated in machine learning, pattern recognition and artificial intelligence. Classification has numerous applications including medical diagnosis, credit evaluation and selective marketing. In recent years, new classification methods have been developed with the new technologies came forth. This paper concludes the new developments of classification method.

**Keywords** Classification, Data mining, Artificial intelligence, Machine learning

## 1 引言

分类是一个比较古老的问题,它在机器学习、模式识别、数据挖掘和人工智能等领域都得到了广泛研究。分类问题可以定义如下:给定一个数据集,所给定的数据集称为训练数据集<sup>[1, 2]</sup>。训练数据集由一组数据库元组(常称作训练样本、实例或对象)构成,每个训练样本是一个由属性值或特征值组成的特征向量,而且每个训练样本还有一个类标号属性。一个具体的样本形式可为:( $v_1, v_2, \dots, v_n; c$ );其中  $v_i$  表示属性值, $c$  表示类标号。给定的训练数据集用来建立一个分类函数(常常也称作分类模型或分类器<sup>[3]</sup>),所建立的分类器用来预测数据库中类标号未知的数据元组的类别。

作为一个古老的问题,分类在很多领域都得到了广泛的研究。到目前为止,已经研究出的经典分类方法主要包括:决策树方法(经典的决策树算法主要包括:ID3 算法<sup>[4]</sup>、C4.5 算法<sup>[5]</sup>和 CART 算法<sup>[6]</sup>等)、神经网络方法(BP 算法<sup>[7]</sup>)、遗传算法(GABIL 系统<sup>[8]</sup>)、贝叶斯分类<sup>[9]</sup>、K-近邻算法<sup>[10]</sup>和基于案例的推理<sup>[11]</sup>。粗糙集方法<sup>[12]</sup>、模糊集方法<sup>[13]</sup>和支持向量机<sup>[14]</sup>是较新的分类方法。

不同的分类方法有不同的特点。用来比较和评估分类方法的标准<sup>[1]</sup>主要有:(1)预测的准确率。模型正确地预测数据类标号的能力;(2)速度。产生和使用模型的计算花费;(3)强壮性。模型对噪声数据或空缺值数据,正确预测的能力;(4)可伸缩性。对于数据量很大的数据集,有效构造模型的能力;(5)可解释性。模型提供的理解和洞察的层次。文<sup>[15]</sup>从预测准确度、模型复杂性以及模型训练时间 3 个方面对 33 种分

类算法进行了比较。另外,分类的效果一般和数据的特点有关,有的数据噪声大,有的有缺失值、有的分布稀疏,有的属性是连续的,而有的则是离散的或混合式的。目前普遍认为不存在某种方法能适合各种特点的数据。

经典分类方法在很多领域都得到了成功的应用,比如决策树方法已经成功地应用到医学诊断、贷款风险评估等领域;神经网络则因为对噪声数据有很好的承受能力而在实际问题中得到了非常成功的应用,比如识别手写字符、语音识别和人脸识别等。但是由于每一种方法都有缺陷,再加上实际问题的复杂性和数据的多样性,使得无论哪一种方法都只能解决某一类问题。近年来,随着人工智能、机器学习、模式识别和数据挖掘等领域中传统方法的不断发展以及各种新方法和新技术的不断涌现,分类方法得到了长足的发展,本文第 2 节将详细介绍这些发展。

## 2 分类方法的新发展

### 2.1 基于粒度计算的分类方法

粒度计算是人工智能领域中新兴起的一个研究方向,是目前人们关注的热点之一。粒度计算是模糊信息粒度理论、粗糙集理论和区间计算理论的超集,是粒度数学的子集,它像一把大伞覆盖了所有有关粒度的理论、方法论、技术和工具的研究<sup>[16]</sup>。关于用粒度计算来解决分类问题,J. T. Yao 和 Y. Y. Yao 进行了一系列的研究<sup>[17~19]</sup>。文<sup>[17]</sup>从粒度计算的角度对两个经典的分类算法 ID3<sup>[4]</sup>和 PRISM<sup>[20]</sup>进行了研究,将它们扩展为基于粒度计算的分类算法;文<sup>[18~19]</sup>比较系统和形式化地研究了以粒度为中心进行分类规则发现的策略,

<sup>\*</sup> 国家自然科学基金资助项目(60573057)。张丽娟 博士研究生,主要研究方向为数据库、数据挖掘;李舟军 教授,博士生导师,主要研究方向为数据库、数据挖掘、进程代数理论、安全协议的形式化验证。

文[18]中使用粒度网络的概念来表达分类知识,并说明了从粒度网络所得到的分类规则比从经典决策树得到的分类规则简短。另外,Yan Zhao 在文[21]中基于粒度计算研究分类问题,通过修改现有的 PRISM 算法来粗化分类规则,使得到的分类规则更容易理解,同时获得了较高的预测精度。

作为一个新的研究方向,虽然目前粒度计算还不成熟,尤其是对粒度计算语义的研究还相当少,但是我们相信随着粒度计算理论本身的不断完善和发展,以及它与其他诸如神经网络、进化计算等软计算方法的互相结合与渗透,在今后几年内它必将在数据挖掘及其它相关领域得到广泛的应用。

## 2.2 基于群的分类方法 (swarm-based approaches)

这种方法可以看作是进化计算的一个新的分支,它模拟了生物界中蚁群、鱼群和鸟群在觅食或者逃避敌人时的行为。纵观文献中对基于群的分类方法的研究,我们将这种方法分为两类:一类是蚁群算法或者蚁群优化 (ant colony optimization ACO),另一类称为 Particle Swarm Optimisers (PSO)。

用蚁群优化来进行分类规则挖掘的算法称为 Ant-Miner<sup>[22,23]</sup>,Ant-Miner 是将数据挖掘的概念和原理与生物界中蚁群的行为结合起来形成的新算法。文[22,23]通过试验比较了 Ant-Miner 与著名的分类算法 CN2<sup>[24]</sup>,结果证明在预测精度方面 Ant-Miner 可与 CN2 相媲美,而 Ant-Miner 发现的规则比 CN2 发现的规则简洁。Ant-Miner3<sup>[25]</sup>对 Ant-Miner 进行了改进,它的预测精度高于 Ant-Miner。

PSO 是进化计算的一个新的分支,它模拟了鱼群或鸟群的行为。PSO 将群中的个体称为 Particles,整个群称为 swarm。在优化领域,PSO 可以与遗传算法相媲美。文[26]将 PSO 用于分类,文章对 Discrete PSO (DPSO)<sup>[27]</sup>、Linear Decreasing Weight PSO (LDWPSO)<sup>[28]</sup> 和 Constricted PSO (CPSO)<sup>[29]</sup> 进行了比较,并选取 CPSO 作为挖掘分类规则的工具。文[30]对 CPSO 进行了改进,并与遗传算法 (GA) 以及 C4.5 算法进行了比较,试验结果表明,在预测精度和运行速度方面 PSO 方法都占优势。

对 ACO 或者 PSO 在数据挖掘中应用的研究仍处于早期阶段,要将这些方法用到实际的大规模数据挖掘中还需要做大量的研究工作。

## 2.3 基于关联规则挖掘的分类方法

关联规则挖掘是数据挖掘中一个重要的研究领域。近年来,对于如何将关联规则挖掘用于分类问题,学者们进行了广泛的研究,取得了一系列研究成果<sup>[31~41]</sup>。纵观这一领域的研究工作,我们大致将其分为两大类:一类是关联分类 (associative classification),另一类是基于 EP (emerging patterns) 或 JEP (jumping emerging patterns) 的分类。

关联分类方法挖掘形如  $\text{condset} \rightarrow C$  的规则,其中 condset 是项(或属性-值对)的集合,而 C 是类标号,这种形式的规则称为类关联规则<sup>[31]</sup> (class association rules CARs)。关联分类方法一般由两步组成:第一步用关联规则挖掘算法从训练数据集中挖掘出所有满足指定支持度和置信度<sup>[1]</sup> 的类关联规则;第二步使用启发式方法从挖掘出的类关联规则中挑选出一组高质量的规则用于分类。属于关联分类的算法主要包括 CBA<sup>[31]</sup>、ADT<sup>[32]</sup>、CMAR<sup>[33]</sup> 等。在提取类关联规则时,CBA 和 ADT 采用了 Apriori 算法<sup>[1]</sup>,而 CMAR 采用了 FP-growth<sup>[1]</sup>;在决定一个新样本的类别时,CBA 和 ADT 简单地挑选置信度最高的规则,即依赖于单一的规则进行分类,而 CMAR 则选择一组高置信度、高度相关的规则,并分析这些

规则之间的相关性,采用一种称为 weighted  $\chi^2$  的新技术进行分类,从而避免了过度拟合数据的问题。关联分类方法的第一步往往产生大量的类关联规则,这对于存储和检索是个极大的挑战。CMAR 采用了一种新颖的数据结构 CR-tree 来存储挖掘出的类关联规则,从而提高了精确度和有效性。另外,文[34]采用关联规则的另一个度量 intensity of implication 对 CBA 算法做了改进,并用试验证明了改进的 CBA 算法比 C4.5、CART 以及最初的 CBA 算法能产生更好的分类结果。文[35]所提出的 PatMat 算法与 CBA 基本思想类似,但是比 CBA 方法简洁、容易理解。CPAR<sup>[36]</sup> (classification based on predictive association rules) 方法结合了关联分类方法和传统分类方法的优点,既有传统方法分类的快速性,又有关联分类方法分类的高精确性。与其它关联分类方法相比,CPAR 方法没有产生大量候选规则,而是采用一种贪婪算法,直接从训练数据集产生一组高质量的类关联规则。在预测新实例的类标号时,CPAR 使用新实例所满足的最好的  $k$  个规则,而不是使用单一规则,从而避免了过度拟合数据的问题。

EP (emerging patterns) 概念是在文[37]中提出的。粗略地讲,EP 是一个项的集合,其支持度由一个类到另一个类显著增加。由于支持度随着类别的不同而发生明显改变,所以 EPs 具有很强的区分能力。JEP (jumping emerging patterns) 是一种特殊的 EP,其支持度由在一个数据集中的零陡峭地增长到另一个数据集中的非零。基于 EP 的分类算法有 CA-EP<sup>[38]</sup> 和 DeEP<sup>[39]</sup>,基于 JEP 的算法有 JEP-Classifier<sup>[40]</sup>,文[40]对 JEP-Classifier 和 CAEP 分类器进行了非常全面的比较。就预测精度来说,对于大规模高维数据集而言,JEP-Classifier 优于 CAEP;对于小规模数据集而言,CAEP 优于 JEP-Classifier,二者互为补充。与 CAEP 和 JEP-Classifier 相比,DeEP 是一种消极的基于实例的分类算法,它的优势在于:能查明与一个待分类实例相关的所有 EP,因而有很高的精确度。DeEP 采用了一系列数据约简技术和简洁的数据表达方法,所以很高效;不需要重新训练分类器就能处理新的训练数据。实验结果表明,DeEP 具有很高的分类精确度、较快的分类速度和较好的伸缩性。另外,文[41]提出了一种特殊类型的 EP: essential jumping emerging patterns (eJEPs)。eJEPs 被认为是具有最好区分能力的高质量模式,因而可用来建立高精度的分类器。文章提出了一种有效挖掘 eJEPs 的单遍扫描算法(将训练数据集压缩为 P-tree 结构,然后基于 P-tree 结构挖掘 eJEPs)。基于 eJEPs 的分类器比 JEP-Classifier 使用少得多的 JEPs,因而更高效。试验证明,在预测精度和训练时间方面,基于 eJEPs 的分类器优于 JEP-Classifier。

比较这些基于关联规则的分类算法,我们得出如下结论:在预测精确度方面,DeEP 优于 JEP-Classifier 和 CAEP,而 JEP-Classifier 和 CAEP 优于 CBA,CBA 又比传统的 C4.5 算法好。

## 2.4 基于 LM (lattice machine) 模型的分类方法

粗略地讲,lattice machine<sup>[42]</sup> (LM) 模型是对数据集的近似,它概括地表达了原数据集的信息,同时和原数据集保持一致,并且不会概括原数据集之外的任何信息。Hui Wang 等人对 LM 模型在分类问题中的应用进行了一系列研究<sup>[42~45]</sup>。其中文[42]针对基于案例的推理 (CBR) 中如何自动设计案例库的问题,提出了基于 LM 的模型。LM 模型将积极学习的强有力性和消极学习推迟处理的灵活性结合起来,自动地发现数据库中的抽象案例 (abstract cases 或者 hypertuples)。

文中提出了一个有效的算法来检索这样的案例。该文的主要贡献在于:所提出的 LM 模型不要求领域知识就能自动地发现给定数据集中的抽象案例。文[43]用基于 LM 的算法进行分类过滤(classificatory filtering),即在约简数据的同时保持分类信息,并提出了一个分类算法 C2,用 C2 在过滤后的数据上进行分类。文章通过统计测试方法证明 LM/C2 算法和 C4.5 算法性能相当。文[44,45]报告了一个新的分类方法 LM/Dens,以提高基于 LM 模型分类的精确度。该方法基于 LM 模型,使用训练数据集的 LM 模型,通过最大化模型的密度来分类新数据。LM/Dens 优于 LM/C2,并且可以与 C5.0 相媲美。文[45]还将 LM/Dens 成功地用于股票市场的数据挖掘。

与很多传统的分类方法(比如决策树归纳)不同,LM 模型旨在近似数据,而决策树归纳则旨在划分数据。

## 2.5 K-近邻算法(K-NN)的新发展

K-NN 是一种有效的分类方法,但是它有两个最大的缺陷:第一,由于要存储所有的训练实例,所以对大规模数据集进行分类是低效的;第二,K-NN 分类的效果在很大程度上依赖于  $k$  值选择的好坏。Gongde Guo 和 Hui Wang 等人<sup>[46-47]</sup>针对 K-NN 的两个缺陷,提出了一种新颖的 KNN 类型的分类方法,称为基于 KNN 模型分类方法。新方法构造数据集的 KNN 模型,以此代替原数据集作为分类的基础,而且新方法中  $k$  值根据不同的数据集自动选择,这样减少了对  $k$  值的依赖,提高了分类速度和精确度。实验证明,基于 KNN 模型的方法在分类精确度上与 C5.0 和标准的 K-NN 相当。文[47]还将这种基于 KNN 模型的方法成功用于文本分类。另外,针对 K-NN 方法的第一个缺陷,Nong Ye and Xiangyang Li 将聚类方法和经典的 K-NN 方法结合起来,提出了一种新颖的分类方法,称为 CCA-S<sup>[48]</sup>。CCA-S 能够处理大规模数据集,可伸缩性好,并且支持增量式学习。但 CCA-S 只能处理连续属性,而且只针对类别为两类的分类问题。如何扩展 CCA-S,以使其能够处理多类别的问题,还有待进一步研究。文[49]将遗传算法和 KNN-Fuzzy 方法<sup>[50]</sup>结合起来,用遗传算法来寻找最优的  $k$  值,从而优化 KNN-Fuzzy 方法,提高了分类精确度。文[51]基于模糊粗集理论提出了一种新的 KNN 分类方法,与传统的 NN 和 fuzzy NN 相比,新方法有更高的预测精度。

## 2.6 支持向量机(SVM)方法的新发展

SVM 是进行分类、聚类和时序分析的有效数据挖掘工具。但是,由于 SVM 的训练时间会随着数据集的增大而增加,所以在处理大规模数据集时,SVM 往往需要较长的训练时间。而实际的数据挖掘应用往往包含了数以百万计的数据,这使得 SVM 很难发挥作用。针对这个问题,文[52,53]用选择性采样或者主动学习方法来训练 SVM,它的基本思想是从整个训练数据集中选择一小部分最有代表性的数据来最大化学习效果。与主动学习思想类似的方法还有随机采样<sup>[54,55]</sup>。这两种方法都需要对数据集进行多遍扫描。与主动学习和随机采样不同,文[56]将层次聚类用于 SVM,以加快 SVM 对大规模数据的处理速度。文中所提出的新方法称为 Clustering-Based SVM(CB-SVM),CB-SVM 用一个层次微聚类(hierarchical micro-clustering)算法对整个数据集进行单遍扫描,为 SVM 提供带有整个数据集统计概要信息的高质量样本。CB-SVM 对于大规模数据集有很好的伸缩性,并且有较高的分类精确度。另外,为了解决实际应用中数据集

大小动态变化的问题,Fung G. 和 Mangasarian O. L.<sup>[57]</sup>提出了增量式 SVM,新提出的方法用于二分类问题。文[58]在文[57]的基础上,提出了一种有效的基于内存的增量算法,以支持多分类问题。

## 2.7 多分类器融合(fusion)的方法

实际应用的复杂性和数据的多样性往往使得单一的分类方法不够有效。因此,学者们对多种分类方法的融合(fusion)进行了广泛的研究,取得了一系列研究成果<sup>[59-76]</sup>。纵观文献中的研究,可以大致将多分类器的融合技术分为以下几类:投票机制(voting)、行为知识空间方法(Behavior-Knowledge Space BKS)、证据理论(Dempster-Shafer theory)、贝叶斯方法和遗传编程(genetic programming GP)。

采用投票机制的方法主要有装袋(bagging<sup>[59]</sup>)和推进(boosting<sup>[60]</sup>)。近一两年来,学者们对 bagging 和 boosting 进行了深入的研究和发展:Buhlmann P. 和 Yu B<sup>[61]</sup>对 bagging 进行了深入的分析;Hothorn T. and Lausen B. 在文[62,63]中将 bagging 用于决策树,并在文[64]中对 bagging 进行了发展,提出了 Double-bagging,采用 bootstrap aggregation 来融合分类器。而 Schwenk 和 Bengio 则将 boosting 用于神经网络,提出了 AdaBoost 方法<sup>[65]</sup>,从而提高了神经网络的预测精度。Peter Buhlmann 和 Bin Yu 提出了 boosting 的一个新变种  $L_2$  Boost<sup>[66]</sup>, $L_2$  Boost 计算简单,且性能可与其他基于 boosting 的方法相媲美。文[67]采用 BKS 进行分类器融合。文[68]用证据理论将 4 个不同的分类方法(SVM, KNN, KNN model-based approach 和 Rocchio)结合起来,形成融合的分类器。用贝叶斯方法进行分类器融合有两种情况:一种是有独立性假设的贝叶斯方法<sup>[69]</sup>,另一种是没有独立性假设的贝叶斯方法<sup>[70-71]</sup>。文[70]把用 K 阶依赖(k-order dependency)表示的离散概率分布的最优近似用于多分类器的融合,用 1 阶依赖和 2 阶依赖发现最优近似。文[71]对 2 阶依赖方法进行了扩展,用 3 阶依赖近似离散概率分布和多分类器融合。文章还通过实验证明了这种基于依赖的贝叶斯方法比基于 BKS 的方法性能好。Langdon 等人对基于遗传编程的分类器融合技术进行了一系列的研究<sup>[72-76]</sup>,其中文[72-74]主要研究了同类型分类器的融合(比如多个神经网络分类器的融合)。文[73]用遗传编程来融合多个神经网络分类器,提高了预测精度;文[74]则进一步证明了基于遗传编程融合多个神经网络的方法比 AdaBoost 方法性能更高。文[75,76]主要研究了不同类型的分类器的融合(比如将神经网络和决策树进行融合)。

除了以上所归纳的 7 个新的发展之外,还有一些方法利用遗传算法或者粗糙集来改进传统分类方法,比如文[77]使用遗传算法直接优化决策树,所提出的系统称为 GATree。与通常的贪婪启发式方法不同,GATree 产生一个动态的、小偏置的、精确的树。文[78]基于粗糙集理论,提出了一种新的属性选择方法,用来构造决策树。文章和传统的基于信息熵的方法进行了比较,证明基于粗糙集的属性选择方法是可行的。文[79]则利用遗传算法来提高粗糙集分类的效果。

**总结** 纵观分类方法近年来的发展,我们可以看出 3 大趋势:一是新分类方法不断涌现,比如基于群分类方法和基于粒度计算的分类方法。新分类方法出现得益于人工智能、机器学习、进化计算和粒度计算等领域中新技术的涌现和发展。二是传统分类方法的进一步发展,比如支持向量机的不断改进和 K-NN 方法的发展。传统分类方法的发展主要利

用了机器学习、进化计算、数据挖掘、模糊集和粗糙集等理论中的原理和方法。三是根据实际问题需要,有针对性地综合众多领域的技术,以提高分类的性能。总之,分类方法向着更加高级、更加综合化和更加多样化的方向发展。

人工智能、机器学习、数据挖掘、进化计算、模糊集和粗糙集等领域的研究成果都对分类方法的发展做出了巨大贡献。相信随着这些领域中相关理论的发展、完善和相互渗透,分类方法也将得到更进一步的发展。

### 参 考 文 献

- 1 Han Jiawei, Kamber M. DATA MINING — Concepts and Techniques [M]. 北京:高等教育出版社,2001
- 2 史忠值. 知识发现[M]. 北京:清华大学出版社,2002
- 3 曾华军,张银奎. 机器学习[M]. 北京:机械工业出版社,2003
- 4 Quinlan J R. Induction of decision trees. *Machine Learning*, 1986, 1;81~106
- 5 Quinlan J R. C4. 5: Programs for Machine Learning. San Mateo, CA; Morgan Kaufmann, 1993
- 6 Breiman L, Friedman J, Olshen R, et al. Classification and Regression Trees. Monterey, CA; Wadsworth International Group, 1984
- 7 Rumelhart D E, Hinton G E, Williams R J. Learning internal representations by error propagation. Rumelhart D E, McClelland J L, editors. *Parallel Distributed Processing*. Cambridge, MA; MIT Press, 1986
- 8 DeJong K A, Spears W M, Gordon D F. Using genetic algorithms for concept learning. *Machine Learning*, 1993, 13; 161~188
- 9 Langley P, Iba W, Thompson K. An analysis of Bayesian classifiers. *AAAI(1990)*, 1990. 223~228
- 10 Cover T M, Hart P E. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 1967, 13;21~27
- 11 Aamodt A, Plazas E. Case-based reasoning; Foundational issues, methodological variations, and system approaches. *AI Comm*, 1994, 7; 39~52
- 12 Pawlak Z. Rough Classification. *Int J Man-Machine Studies*, 1984,20; 469~483
- 13 Zadeh L A. Fuzzy sets. *Information and Control*, 1965, 8; 338~353
- 14 Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines. Cambridge University Press, 2000
- 15 Lim T, Loh W, Shih Y. A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-three Old and New Classification Algorithms. *Machine Learning*, 2000, 40; 203~228
- 16 Yao Y Y. Granular computing; basic issues and possible solutions. In: *Proceedings of the 5th Joint Conference on Information Sciences*, 2000. 186~189
- 17 Yao J T, Yao Y Y. A granular computing approach to machine learning. <http://www2.cs.uregina.ca/~jtyao/Papers/GrcMining-1534.pdf>, 2002
- 18 Yao Y Y, Yao J T. Induction of classification rules by granular computing. In: *Proceedings of The Third International Conference on Rough Sets and Current Trends in Computing*, 2002
- 19 Yao Y Y, Yao J T. Granular computing as a basis for consistent classification problems. In: *Proceedings of PAKDD'02 Workshop on Toward the Foundation of Data Mining*, 2002. 101~106
- 20 Cendrowska J. PRISM; an algorithm for inducing modular rules. *International Journal of Man-Machine Studies*, 1987, 27; 349~370
- 21 Zhao Yan. Coarsening classification rules on basis of granular computing. In: *Taw A Y, Goodwin S D. Eds. Canadian AI 2004*, LNAI 3060, 2004. 578~579
- 22 Parpinelli R S, Lopes H S, Freitas A A. Data Mining with an Ant Colony Optimization Algorithm. *IEEE Trans. on Evolutionary Computation*, 2002, special issue on Ant Colony algorithms
- 23 Parpinelli R S, Lopes H S, Freitas A A. Mining Comprehensible Rules from Data with an Ant Colony Algorithm. In: *Bittencourt G, Ramalho . eds. SBIA 2002, LNAI 2507*, 2002. 259~269
- 24 Clark P, Niblett T. The CN2 induction algorithm. *Machine Learning*, 1989. 261~283
- 25 Liu Bo, Abbass A, McKay B. Classification rule discovery with ant colony optimization. *IEEE Computational Intelligence Bulletin*, 2004, 3(1);31~35
- 26 Sousa T, Silva A, Neves A. Particle Swarm Optimisation as a New Tool for Data Mining. *International Parallel and Distributed Processing Symposium(IPDPS)*. Nice, France,2003
- 27 Kennedy J, Eberhart R C. *Swarm Intelligence*. Morgan Kaufmann,2001
- 28 Shi Y, Eberhart R C. Empirical Study of Particle Swarm Optimization. In: *Proceedings of the 1999 Congress of Evolutionary Computation*. Piscatay,1999
- 29 Clerc M, Kennedy J. The Particle Swarm-Explosion, In: *Stability and Convergence in a Multidimensional Complex Space*. *IEEE Transactions on Evolutionary Computation*, 2002, 6(1)
- 30 Sousa T, Silva A, Neves A. A particle swarm data miner. In: *Pires F M P, Abreu S,eds. EPIA 2003, LNAI 2902*, 2003. 43~53
- 31 Liu B, Hsu W, Ma Y. Integrating Classification and Association Rule Mining. In: *Proc. of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, New York, 1998. 80~86
- 32 Wang K, Zhou S, He Y. Growing decision tree on support-less association rules. In: *Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'00)*, Boston, 2000. 265~269
- 33 Li W, Han J, Pei J. CMAR: Accurate and Efficient Classification Based on Multiple Class-Association Rules. In: *Proc. of the 1st IEEE International Conference on Data Mining(ICDM 2001)*, San Jose, California, 2001. 369~376
- 34 Janssens D, Wets G, Brijs T, et al. Integrating Classification and Association Rules by proposing adaptations to the CBA Algorithm. <http://citeseer.ist.psu.edu/janssens03integrating.html>. 2003
- 35 Pijls W H L M, Potharst R. Classification and target group selection based upon frequent patterns. <http://ideas.repec.org/p/dgr/eureri/200043.html>. 2004
- 36 Yin Xiaoxin, Han Jiawei. CPAR: Classification based on Predictive Association Rules. in *SIAM International Conference on Data Mining(SDM'03)*, San Francisco, CA, May 2003
- 37 Dong G, Li J. Efficient mining of emerging patterns; Discovering trends and differences. In: *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining*, San Diego, USA(SIGKDD'99), 1999. 43~52
- 38 Dong G, Zhang X, Wong L, et al. CAEP: Classification by aggregating emerging patterns. In: *Proc. of the Second International Conference on Discovery Science*, Tokyo, Japan, 1999. 30~42
- 39 Li J, Dong G, Ramamohanarao K. Instance-Based Classification by Emerging Patterns. *PKDD*, 2000. 191~200
- 40 Li J, Dong G, Ramamohanarao K. JEP-Classier: Classification by Aggregating Jumping Emerging Patterns. *Knowledge and Information Systems*, 2001,3(2);131~145
- 41 Fan Hongjian, Kotagiri R. An efficient single-scan algorithm for mining essential jumping emerging patterns for classification. *Chen M S, Yu P S, Liu B. Eds. PAKDD 2002, LNAI 2336*,

2002. 456~462
- 42 Wang H, Dubitzky W, Düntsch I, et al. Bell. A lattice machine approach to automated casebase design: Marrying lazy and eager learning. In: Proc. IJCAI99, Stockholm, Sweden, 1999. 254~259
- 43 Wang H, Düntsch I, Gediga G. Classificatory filtering in decision systems. *International Journal of Approximate Reasoning*, 2000, 23:111~136
- 44 Wang Hui, Düntsch I, Bell D, et al. Classification through maximizing density. In: Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM. 01), 2001. 655~656
- 45 Wang Hui, Bell D, Düntsch I. A Density Based Approach to Classification. In: SAC2003, 2003. 470~474
- 46 Guo Gongde, Wang Hui, Bell D A, et al. KNN Model-Based Approach in Classification. In: CoopIS/DOA/ODBASE 2003, 2003. 986~996
- 47 Guo Gongde, Wang Hui, Bell D A, et al. A kNN Model-Based Approach and Its Application in Text Categorization. In: CICALING 2004, 2004. 559~570
- 48 Ye Nong, Li Xiangyang. A machine learning algorithm based on supervised clustering and classification. In: Liu J, et al, eds. AMT 2001, LNCS 2252, 2001. 327~334
- 49 Rosa L A, Ebecken N F F. Data mining for data classification based on the KNN-Fuzzy method supported by genetic algorithm. In: VECPAR 2002, LNCS 2565, 2003. 126~133
- 50 Keller J M, Gray M R, Givens J A Jr. A Fuzzy K-Nearest Neighbor Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics*, 1985, SMC-15(4):258~260
- 51 Bian H Y. Fuzzy-Rough Nearest Neighbor Classification: an Integrated Framework. In: Proceedings of IASTED International Symposium on Artificial intelligence and Applications, 2002. 160~164
- 52 Schohn G, Cohn D. Less : Active learning with support vector machines. In: Proc. 17<sup>th</sup> Int Conf. Machine Learning, Stanford, CA, 2000
- 53 Tong S, Koller D. Support vector machine active learning with applications to text classification. In: Proc. 17<sup>th</sup> Int Conf. Machine Learning, Stanford, CA, 2000
- 54 Balczar O W J L, Dai Y. A random sampling technique for training support vector machines. In: The 2001 IEEE Int Conf. Data Mining, San Jose, CA, 2001
- 55 Lee Y J, Mangasarian O L. RSVM: Reduced support vector machines. First SIAM Int Conf Data Mining, Chicago, IL, 2001
- 56 Yu Hwanjo, Yang Jiong, Han Jiawei. Classifying Large Data Sets Using SVMs with Hierarchical Clusters. SIGKDD' 03, Washington, DC, USA, Aug. 2003
- 57 Fung G, Mangasarian O L. Incremental support vector machine classification. In: Grossman R, Mannila H, Motwani R, eds. Proceedings of the Second SIAM International Conference on Data Mining, SIAM(2002), 2002. 247~260
- 58 Tveit A, Hetland M L. Multicategory Incremental Proximal Support Vector Classifiers. In: Palade V, Howlett R J, Jain L C, eds. KES 2003, LNAI 2773, 2003. 386~392
- 59 Breiman L. Bagging predictors. *Machine Learning*, 1996, 24:123~140
- 60 Schapire F, Freund Y, Schapire R E. Experiments with a new boosting algorithm. In: Machine Learning: Proceedings of the thirteenth International Conference, Morgan Kaufmann, 1996. 148~156
- 61 Buhlmann P, Yu B. Analyzing bagging. *The Annals of Statistics*, 2002, 30:927~961
- 62 Hothorn T, Lausen B. Building classifiers by bagging trees. Preprint, Friedrich-Alexander-University Erlangen-Nuremberg, URL. <http://www.mathpreprints.com/>
- 63 Hothorn T, Lausen B, Benner A, et al. Bagging survival trees. *Statistics in Medicine*, 2004, 23: 77~91
- 64 Hothorn T, Lausen B. Double-bagging: Combining classifiers by bootstrap aggregation. *Pattern Recognition*, 2003, 36: 1303~1309
- 65 Schwenk H, Bengio Y. Boosting neural networks. *Neural Computation*, 2000, 12(8):1869~1887
- 66 Buhlmann P, Yu Bin. Boosting with the  $L_2$ -Loss: Regression and Classification. *Journal of the American Statistical Association*, 2003, 98: 324~338
- 67 Huang Y S, Suen C Y. A Method of Combining Multiple Experts for the Recognition of Unconstrained Handwritten Numerals. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1995, 17(1):90~94
- 68 Bi Yaxin, Bell D A, Wang Hui, et al. Combining Multiple Classifiers Using Dempster's Rule of Combination for Text Categorization. In: MDAI2004, 2004. 127~138
- 69 Xu L, Krzyzak A, Suen C Y. Several Methods for Combining Multiple Classifiers and Their Applications in Handwritten Character Recognition. *IEEE Trans on System, Man and Cybernetics*, 1992, 22(3): 418~435
- 70 Kang H J, Kim K, Kim J H. Optimal Approximation of Discrete Probability Distribution with kth-order Dependency and Its Applications to Combining Multiple Classifiers. *PRL*, 1997, 18(6): 515~523
- 71 Kang Hee-Joong, Doermann D. Combining Multiple Classifiers Based on Third-Order Dependency. In: Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003)
- 72 Langdon W B, Buxton B F. Genetic programming for combining classifiers. In: Spector L, et al., eds. Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001). San Francisco, California, USA, July 2001. 66~73
- 73 Langdon W B, Barrett S J, Buxton B F. Genetic programming for combining neural networks for drug discovery. In: Roy R, et al. eds. Soft Computing and Industry Recent Applications. Springer-Verlag, September 2001. 597~608
- 74 Langdon W B, Barrett S J, Buxton B F. Comparison of AdaBoost and Genetic Programming for combining Neural Networks for Drug Discovery. Applications of Evolutionary Computing. In: Raidl G R, et al., eds. LNCS 2611, Essex, UK: Springer-Verlag, April 2003. 87~98
- 75 Langdon W B, Buxton B F. Genetic programming for improved receiver operating characteristics. In: Kittler J, Roli F, eds. Second International Conference on Multiple Classifier System, vol 2096 of LNCS. Cambridge, Springer Verlag, July 2001. 68~77
- 76 Langdon W B, Barrett S J, Buxton B F. Combining decision trees and neural networks for drug discovery. In: Foster J A, et al., eds. Genetic Programming. Proceedings of the 5<sup>th</sup> European Conference, EuroGP 2002, vol 2278 of LNCS. Kinsale, Ireland: Springer-Verlag, April 2002. 60~70
- 77 Papagelis A, Kalles D. Breeding decision trees using evolutionary techniques. In: Proc. 18<sup>th</sup> Int Conf on Machine Learning, San Mateo: Morgan Kaufmann, 2001. 393~400
- 78 Wei J M. Rough set based approach to selection of node. *International Journal of Computational Cognition*, 2002, 1(2):25~40
- 79 Lingras P. Unsupervised Rough Set Classification Using GAs. *Journal of Intelligent Information Systems*, 2001, 16: 215~228