基于最大熵原理的空间特征选择方法*

宋国杰[1+], 唐世渭[1,2], 杨冬青[1], 王腾蛟[1,2]

[1](北京大学 计算机科学技术系,北京 100871)

[2](北京大学 视觉与听觉信息处理国家重点实验室,北京 100871)

A Spatial Feature Selection Method Based on Maximum Entropy Theory

SONG Guo-Jie[1+], TANG Shi-Wei[1,2], YANG Dong-Qing[1], WANG Teng-Jiao[1,2]

[1](Department of Computer Science and Technology, Peking University, Beijing 100871, China)

[2](National Laboratory on Machine Perception, Peking University, Beijing 100871, China)

+ Corresponding author: Phn: 86-10-62763510, E-mail: sgj@db.pku.edu.cn

http://db.cs.pku.edu.cn

Received 2002-08-09; Accepted 2002-12-23

Song GJ, Tang SW, Yang DQ, Wang TJ. A spatial feature selection method based on maximum entropy theory. *Journal of Software*, 2003,14(9):1544~1550.

http://www.jos.org.cn/1000-9825/14/1544.htm

Abstract: Feature selection has an important application in the field of pattern recognition and data mining etc. However, in real world domains, if there are spatial data operated in the application, the performance of feature selection will be decreased because of without considering the characteristic of spatial data. In this paper, a feature selection method from the point of the characteristic of spatial data, named MEFS (maximum entropy feature selection), is proposed. Based on the theory of maximum entropy, MEFS uses mutual information and Z-test technologies, and takes two-step method to execute feature selection. The first step is predicate selection, and the second step is to choose relevant dataset corresponding to each predicate. At last, the experiments between feature selection algorithms MEFS and RELIEF, and between ID3 classification algorithm and classification algorithm based on MEFS are carried out. The experimental results show that the MEFS algorithm not only saves feature selection and classification time, but also improves the quality of classification.

Key words: spatial data mining; spatial feature selection; maximum entropy theory; mutual information; decision tree

摘 要: 特征选择在模式识别和数据挖掘等领域都有十分广泛的应用.然而,当涉及空间数据时,由于传统特征选择方法没有很好地考虑数据的空间特性,所以会导致特征选择结果性能下降.从空间数据本身的特性出发,提出一种特征选择方法 MEFS(maximum entropy feature selection).MEFS 在基于最大熵原理的基础上,运用互信息和 Z-测试技术,采用两步方法进行空间特征选择.第 1 步,空间谓词选择;第 2 步,选择与每个空间谓词对应的相关

* Supported by the Foundation of the Innovation Research Institute of PKU-IBM (北京大学-IBM 中国研究中心联合实验室资助项目); the National Grand Fundamental Research 973 Program of China under Grant No.G1999032705 (国家重点基础研究发展规划(973))

第一作者简介: 宋国杰(1975－),男,河南原阳人,博士生,主要研究领域为数据仓库,数据挖掘.

属性集.最后,分别对 MEFS 方法和 RELIEF 方法以及基于 MEFS 的分类方法与决策树算法 ID3 分别进行了实验比较.实验结果表明,MEFS 方法不仅可以节约特征提取和分类时间,而且也极大地提高了分类质量.

关键词: 空间数据挖掘;空间特征选择;最大熵理论;互信息;决策树

中图法分类号: TP391 **文献标识码**: A

特征选择在模式识别、数据挖掘领域有着十分广泛的应用,同时也是需要有效解决的重要问题.特征选择是指,从已知一组特征集中按照某一准则选择出有很好的区分特性的特征子集,或按照某一准则对特征的分类性能进行排序,用于分类器的优化设计.目前的特征选择方法主要是传统的模式识别方法[1~5],这些方法主要集中讨论了非空间数据特征提取的问题,对空间数据环境下的特征提取问题,文献[9]给出了讨论,但总体而言,研究还不够深入.当应用中涉及空间数据时,由于传统特征选择方法没有很好地考虑空间数据特性,所以导致特征选择结果和性能下降.因此,空间特征选择问题变得十分迫切和必要.

本文从空间数据特性的角度出发,提出一种新的特征选择方法 MEFS(maximum entropy feature selection),并应用到我们研制的空间数据挖掘原型系统 SpatialMiner 中.MEFS 基于最大熵原理,运用互信息和 Z-测试技术,采用两步方法进行空间特征选择:首先是空间谓词选择,然后选择与每个空间谓词对应的相关属性集.最后对 MEFS 方法和 RELIEF[6]方法以及基于 MEFS 的分类方法与 ID3 算法分别进行了实验比较,结果表明,MEFS 方法不仅可以节约特征提取和分类时间,而且也极大地提高了分类质量.

1 最大熵原理

最大熵原理在文献[7,8]中给出了详细的描述,其基本思想是:给定训练样本,选择一个与训练样本一致的模型.最大熵模型应选择与这些观察相一致的概率分布,而对于除此之外的情况,模型赋予均匀的概率分布.

1.1 问题描述

假设特征选择的分类属性值构成随机过程 P 所有输出值 Y.对于每个 $y\in Y$,其出现均受与之相关的决策属性值 x 的影响.已知与 Y 相关的所有决策属性值组成的集合为 X,则模型的目标是:对给定的所有决策属性 $x\in X$,计算输出为 $y\in Y$ 的条件概率,即对 $p(y|x)$进行估计,其中 $y\in Y$ 且 $x\in X$.因此,特征选择的目的就是从众多决策属性中选择出对分类属性具有明显表征作用的特征.

1.2 训练数据

特征选择过程是在抽样数据的基础上,抽样数据来自采样数据库,对空间而言还包含空间数据信息,表示为 $(x_1,y_1),(x_2,y_2),\ldots,(x_i,y_i),\ldots,(x_n,y_n)$.其中,$x_i$ 表示决策属性,或为空间数据,或为非空间数据,y_i 是分类属性,是由专家提供的类标号.在训练数据的基础上,可以用概率分布的极大似然对训练样本进行表示.即

$$\tilde{p}(x,y)\equiv\frac{freq(x,y)}{\sum_{x,y}freq(x,y)},$$

其中 $freq(x,y)$表示(x,y)在样本中出现的次数.

1.3 定 义

定义 1(特征). 设 $x\in X$ 且 $x=w_1w_2\ldots w_n$,设 c 是 x 的子串(长度≥1),若 c 对 $y\in Y$ 具有表征作用,则称(c,y)为模型的一个特征.特征分为原子特征和复合特征.若串 c 的长度为 1,则称(c,y)为原子特征,否则,称(c,y)为复合特征.

定义 2(特征函数). 特征函数是一个二值表征函数,表示(x',y')是否与特征(c,y)有关.定义(x',y')关于特征(c,y)的特征函数为

$$f_{(c,y)}(x',y')=\begin{cases}1, & \text{若}c\text{是}x'\text{的子串,且}y'=y\\ 0, & \text{否则}\end{cases}.$$

定义 3(约束). 设 $\tilde{p}(f)$为特征 f 对于经验概率分布 $\tilde{p}(x,y)$的数学期望,表示为 $\tilde{p}(f)=\sum_{xy}\tilde{p}(x,y)f(x,y)$,$p(f)$为

特征 f 对于由模型确定的概率 $p(x,y)$的数学期望,表示为 $p(x,y)=\sum_{xy} p(x,y)\,f(x,y)$,而 $p(x,y)=p(x)p(y|x)$,令 $p(x)=\tilde{p}(x)$,则限定所求模型的概率为在样本中观察到的事件的概率,而不是所有可能出现的事件的概率.若 f 对模型有用,则令 $p(f)=\tilde{p}(f)$为约束.

1.4 最大熵原理的引入

假设存在 n 个特征 $f_i(i=1,2,\ldots,n)$,则模型属于约束所产生的模型集合,即

$$C=\{p\in P \mid p(f_i)=\tilde{p}(f_i), i\in\{1,2,...,n\}\},$$

而满足约束条件的模型有很多,模型的目标是产生在约束集下具有最均匀分布的模型,而条件概率 $p(y|x)$均匀性的一种数学测量方法为条件熵,定义为

$$H(p)=-\sum_{x,y}\tilde{p}(x)p(y|x)\log p(y|x),$$

其中 $0\le H(p)\le\log|y|$.

最大熵原理. 若在允许的概率分布 C 中选择模型,具有最大熵的模型 $p_*\in C$ 即为所选模型.即

$$p_*=\arg\max_{p\in C} H(p).$$

2 空间特征选择

利用最大熵原理求取空间特征包含特征选择和参数估计.特征选择是选出对分类对象有明显表征作用的属性;参数估计是用最大熵原理对每一个特征进行参数估值,使每个特征对应于一个特征参数.特征参数用来反映决策属性与分类属性之间的关联强度.

本文基于空间数据特性,提出了两步方法进行空间特征选择:谓词提取和相关属性选择.谓词提取选出能够以某种空间谓词(或函数)表征分类对象的数据集.相关属性选择在已选择谓词的基础上,选出依附于该谓词而且能够表征分类对象的非空间属性.

2.1 相关概念

(1) 互信息.互信息是测量搭配强度的一个物理量.若某一变量 x 对 y 有表征意义,则 y 与该 x 的互信息较大.计算如下式:

$$I(x,y)=\log\frac{p(x,y)}{p(x)p(y)}.$$

(2) Z-测试.我们可以求取变量间的关联强度,但如果特征选择直接确定选择互信息大于某一阈值的上下文信息为特征时,则对于不同互信息的分布,阈值也不相同,这样算法难以操作.我们需要一种方法来进行变换,使得所有变量互信息的分布服从统一的准则.Z-测试正是这样的一个测度,它可以将互信息的分布进行标准变换,将其变为标准的正态分布.这样,不论互信息如何进行分布,都可以从一个统一的阈值开始进行求解.计算如下式:

$$z_{xy}=\frac{I(y,x)-E_y}{\sqrt{\mu_y}},$$

其中 E_y 表示互信息均值,表示为 $E_y=\frac{1}{n}\sum_{x\in X} I(y,x)$;$\mu_y$ 表示均方差,表示为 $\mu_y=\frac{1}{n}\sum_{x\in X}(I(y,x)-E_y)^2$.

(3) IIS.建立最大熵模型的关键是要选出具有预期作用的特征,只有这样才能保证得到的解是对模型最有用的解.IIS[9]算法较好地解决了这一问题.算法假设满足最大熵条件的概率 $p(x,y)$具有 Gibbs 分布的形式:

$$p_\lambda(y|x)=\frac{1}{Z_\lambda(x)}\exp\left(\sum_i \lambda_i f_i(x,y)\right),$$

其中

$$Z_\lambda(x)=\sum_y \exp\left(\sum_i \lambda_i f_i(x,y)\right),$$

$Z_\lambda(x)$为归一常量,保证对所有 x, $\sum_y p_\lambda(y|x)=1$.

(4) 模型质量度量.利用特征选择算法得到特征集确定的模型 $p.p$ 与由训练集确定的概率模型 $\tilde{p}$ 之间的距离作为度量所求模型质量的尺度,这里采用相对熵来度量.相对熵被用于衡量两个随机分布的差距,定义为

$$D(\tilde{p}\|p)=\sum_{x,y}\tilde{p}(x,y)\log\frac{\tilde{p}(y|x)}{p(y|x)}.$$

2.2 空间谓词的提取

Koperski 在文献[6]中给出了一种谓词提取方法.提取过程分为两步:首先进行粗略计算,然后在概念层次的基础上只对有希望的模式进行细化计算.简单而言,谓词提取就是发现满足某种谓词关系的、可以用于描述分类对象的对象.但该方法提取的结果都是对象类(如铁路),而不是具体的某类对象(长铁路).但在实际的分类过程中,不同的分类属性可能与不同的某类对象而不是与对象类相关.所以,特征提取不仅应该考察对象类的提取,更重要的是提取能够标识分类对象的某类对象.

2.2.1 谓词提取算法

空间谓词和空间函数作为谓词提取的对象,用来描述分类对象.为了利用最大熵原理进行提取,首先要将空间谓词和函数转化为样本形式 $S=(P_1(x_1),y_1),\ldots,(P_i(x_j),y_k),\ldots,(P_l(x_m),y_n)$.其中,$(P_i(x_j),y_k)$表示分类对象 y_k 与决策对象属性 x_j 满足关系 P_i.求取的最终结果是满足特征参数阈值的谓词集及其参数估计值.具体算法描述如下:

算法 1. **谓词提取(predicate_selection)**.

输入:样本集 S,特征参数阈值 t,质量度量阈值 ε.

输出:特征集 C 和特征参数集 P.

步骤描述如下:

1) for each $(x,y)\in S$ do
2) compute $I(x,y)$
3) if $I(x,y)>>0$ then
4) $C=C\cup(x,y)$
5) compute Z_{xy}
6) if $Z_{xy}>t$ then $C=C\cup(x,y)$
7) $P=P\cup IIS(C)$
8) compute $p_\lambda(y|x)$
9) compute $D(\tilde{p}\|p)$
10) if $D(\tilde{p}\|p)\geq\varepsilon$ then
11) $t=t-\Delta t$ return 5)

2.2.2 时间复杂性分析

算法的计算量由互信息、特征参数及相对熵的计算量 3 部分组成.互信息的计算量与空间关系数据集的规模线性相关.假设 $P=|Y|,N=|X|$,则互信息的计算复杂度为 $O(P\times N)$.由于特征求取和参数估计的过程是一个迭代的过程,而且由互信息的正态分布特性可知,算法必将在有限步内收敛,假设为 k 次循环.假设第 k 次循环中参数估计时间量为 IIS_k(也为最大量,因为随着迭代的进行,特征数目有所增加).同样,相对熵的最大计算量为 D_k,则它们的时间复杂度为 $O(k\times IIS_k\times D_k)$.所以总的时间复杂度为 $O(P\times N)+O(k\times IIS_k\times D_k)$.

2.3 非空间决策属性选择

决策属性包括空间和非空间两种.经过上述谓词提取之后,得到空间决策属性集合 C 及特征参数集 P.从谓词提取结果可知,某些空间决策属性对于所有的分类对象可能都具有较强的关联程度,但这对分类无用,需要一种方法将这些空间决策属性剔除,仅保留对分类具有明显作用的部分.我们引入如下定义:

定义 4(特征强度(feature strength)). 假设 D 是决策属性集,d 是 D 中的一个元素,F 是一个空间分类对象对应的空间决策属性集.$sig^s(d)$表示决策属性 d 在集合 s 中的特征参数和.$card(s)$表示集合 s 中元素的特征参数总和,则决策属性 d 在 F 中相对于 D 的特征参数强度可表示为 $sig_F^D(d)$,定义如下:

$$sig_F^D(d)=\frac{sig^F(d)}{card(F)}\Big/\frac{sig^D(d)}{card(D)}.$$

假设 *significance* 表示特征参数强度阈值,那么满足如下条件的决策属性为具有分类作用的决策属性:

$$sig_F^D \geq significance \text{ 或者 } sig_F^D \leq significance.$$

2.3.1 非空间决策属性选择算法

非空间决策属性依附于某类空间对象,与不同分类对象对应的具有不同谓词关系的空间决策对象具有不同的相关属性.如与某一分类对象具有 close_to 关系的中等发达城市和与具有 contain 关系的发达城市就有不同的非空间决策属性对应.因此,需要对空间决策属性集合 C 中的元素逐个进行考虑,选取与之对应的非空间决策属性(原子属性或者合成属性).在进行非空间决策属性选择前后,需要利用定义 4 所定义的特征强度过滤无用的空间决策属性和非空间决策属性.

算法 2. 相关属性选择(relevent_property_selection)

输入:空间决策属性集 C,特征参数集 P,特征强度阈值 *significance*.

输出:非空间决策属性集 *non_C*,特征参数集 *non_P*.

步骤描述如下:

1) for each $c\in C$ do
2) if $sig_F^D(c)>significance$ then
3) $C'=C'\cup c$; $P'=P'\cup p$
4) for each $c\in C'$ do
5) S=choose_non_spatial_properties(c)
6) $C''(P'')$=call predicate_selection
7) $non\_C'=non\_C'\cup C''$
8) $non\_P'=non\_P'\cup P''$
9) for each $c\in non\_C'$ do
10) if $sig_F^D(c)>significance$ then
11) $non\_C=non\_C\cup c$
12) $non\_P=non\_P\cup p$

注:在本算法中利用算法 1 进行非空间决策属性的选择是合理的,因为转化为关系型之后无差别.

时间复杂性分析.本算法时间复杂性由两部分组成:特征强度和算法 1 的时间复杂性.由于计算特征强度的时间复杂性较低,故忽略不计.算法 1 的复杂性分析在第 2.2.2 节已经给出,这里不再赘述.假设$|C'|=m$,则本算法的时间复杂性为 $O(m\times P\times N)+O(m\times k\times IIS_k\times D_k)$.

2.3.2 结果分析

非空间决策属性选择结果是特征集和与之对应的特征参数.其中特征集不仅包含非空间决策属性,而且也包含产生它的空间决策属性(谓词集).如对于一空间决策属性$(P_i(x_j),y_k)$经过算法 2 中的步骤 5),可以得到如下形式的样本元素:$(P_i(x_j)p_1\ldots p_i\ldots p_n,y_k)$,其中 p_i 或者是原子属性或者是合成属性.因此,非空间决策属性选择的结果不仅代表非空间的意义,还包含了空间决策属性的全部内容.

3 实验结果

3.1 基于空间特征的分类分析

在本文中,我们将基于最大熵理论进行空间特征选择,并将结果根据人均储蓄情况对全国的县进行分类.模型输入:已知全国县的人均储蓄分布及其邻居关系,分类对象 Y 为按人均储蓄分类的县,决策属性为县本身的属性以及与县具有空间谓词关系的对象以及属性.模型输出:影响县人均储蓄的特征集 S 以及表征参数λ.根据特征选择结果,基于抽样数据集,首先进行 RELIEF 特征抽取方法和 MEFS 方法在选择时间上的比较,然后进行基于 MEFS 的分类方法和 ID3 算法在分类时间和质量上的比较.

3.2 实验过程

(1) 实验环境与数据

采用 IBM 公司数据库 DB_2 和空间环境 Spatial Extender 作为数据操作环境,Java 为开发平台.数据取自国家

地理信息系统中国地图数据,实验中涉及的图层有:分类对象是全国 2 500 左右个县,预测对象包括铁路、公路以及对象自身包括面积、人均消费、粮食产量以及收入等属性,分类对象按人均储蓄属性分为 5 个类 t_1~t_5,谓词关系为 contain.从每个分类对象集合中选取 400 个对象用于训练集,余下的 500 个对象用于测试.

(2) 特征选择和参数估计

按本文提出的特征选择算法和参数估计算法,结果见表 1.

Table 1 The results of feature analysis

表 1 特征分析结果

Personal deposit	Feature set S	Feature factor	Spatial feature set S	Feature factor
(0~328)	Avgconsume (lower)	1.019 32	Contain (x,rail)/area (low)	0.599 4
(0~328)	Avgfinnace (lower)	0.931 536	Contain (x,road)/area (low)	0.817 1
(0~328)	Avgcorp (lower)	0.548 812		
(328~512)	Avgconsume (low)	0.529 46	Contain (x,rail)/area (medium)	0.734 1
(328~512)	Avgfinance (low)	0.402 27	Contain (x,road)/area (medium)	0.869 1
(512~800)	Avgsonsume (medium)	0.400 781	Contain (x,rail)/area (medium)	0.765 6
(512~800)	Avgfinace (medium)	0.341 808	Contain (x,road)/area (medium)	1.143 6
(800~1440)	Area (lower)	0.347 563	Contain (x,rail)/area (high)	0.936 3
(800~1440)	Avgconsume (higher)	0.717 677	Contain (x,road)/area (high)	0.554 6
(800~1440)	Avgfinance (higher)	0.580 192		
(800~1440)	Avgcorp (highest)	0.352 865		
(1440~20803)	Area (lowest)	0.641 875	Contain (x,rail)/area (high)	1.087 2
(1440~20803)	Avgconsume (highest)	1.311 04	Contain (x,road)/area (high)	1.068 5
(1440~20803)	Avgfinance (highest)	1.254 38		

由此得出如下结论:县人均储蓄与 contain(x,rail)/area,contain(x,road)/area(单位面积铁路和公路长度)成正比,与县面积成反比,与 avgconsume(人均消费),avgfinance(人均财政收入),avgcorp(人均粮食产量)成正比.

(3) 性能比较

空间分类是空间数据挖掘的一个重要研究分支,ID3 算法是经典的也是高效的算法之一.下面在上述数据集的基础上,采用 ID3 算法并基于本文的 MEFS 方法进行分类比较.比较过程如下:

基于 MEFS 的分类过程描述如下:对于给定的抽样样本集,经过空间特征提取过程可以得到特征和参数集$\langle S,\lambda\rangle$.对于待识别的样本集合的每个元素,用分类特征公式计算各个不同分类对象对应的特征公式的值,最后取特征值最大者对应的类为得到的类.

下面分别对特征选择时间、分类时间和分类准确度进行实验比较,结果如下:

(1) 特征选择时间比较

图 1 是在上述数据集之上对 MEFS 特征选择方法和 RELIEF 方法进行比较之后的时间变化曲线.由图 1 可知,本文的 MEFS 方法在提取过程中没有 RELIEF 算法的效率高.这是因为 RELIEF 算法进行的是模式选择,也就是说它选出的特征是对象和属性类的集合,而不是表征分类对象的某类对象.而本文的 MEFS 方法考察的是表征分类对象的对象和属性类,得出了真正表征分类对象的对象和属性,因此在效率上有所降低.

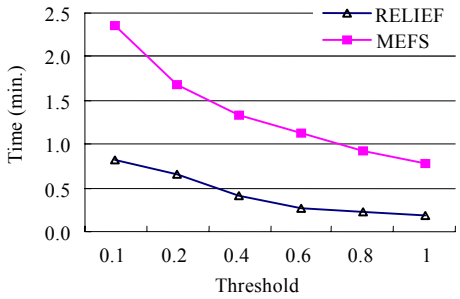

Fig.1 The comparison of feature selection in time

图 1 特征选择时间比较

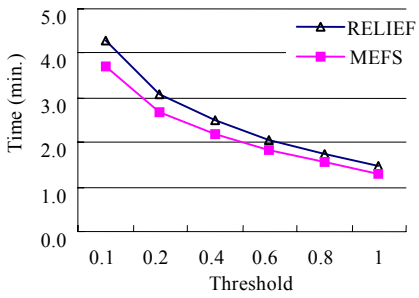

Fig.2 The comparison of total consume time

图 2 总花费时间比较

但是,效率的降低在如下方面得到补偿:(1) 特征选择精度得到提高;(2) 分类过程不需要构造决策树,免去这部分时间的消耗,而且对于一棵大决策树而言,这个消耗也是很大的.因此,如果特征选择时间总和与决策树

的构造时间总和进行比较,可以得到如图 2 所示的性能比较结果.

图 2 的对比说明,就总体特征提取和构造决策树的代价总体而论,MEFS 算法在时间上还是优于 RELIEF 算法的.这是因为 MEFS 方法的分类规则是以分类特征公式而表达的,而 RELIEF 算法必须构造一棵决策树,构造该树的时间耗费是随着特征集规模的大小而呈正比变化的.

(2) 分类时间和质量比较

实验结果表明,采用本文的基于 MEFS 的方法进行空间分类与采用 ID3 决策树算法相比,时间复杂度要低一些.这是因为决策树判定的过程就是与当前分类属性和树的当前分支比较,每进行一个分支,需要扫描一遍当前抽样集合的元组,扫描遍数为当前所沿路径的深度.但是基于 MEFS 的方法只需进行一遍扫描,将当前抽样元组含有的决策特征属性映射到各个分类对象对应的特征公式即可.

由表 2 可知,基于 MEFS 方法的分类的精度也比 ID3 有所提高,这是因为基于 MEFS 的分类方法所基于的分类特征是经过过滤的,每个分类属性都对应于自己本身的特征对象和属性,这样不但提高了分类的准确度和时间消耗,而且也避免了 ID3 在建造决策树时对噪声的干扰.对 ID3 算法而言,如果输入数据不完整或出现噪声,就会严重影响决策树算法的预测准确度.空间数据分类在分类标签中使用的空间谓词往往带有不确定性和不完整信息,数据噪声出现的几率较大,ID3 算法难以克服这些问题.利用 MEFS 方法进行特征提取,可以去掉多余的属性和谓词,提高发现效率,降低错误率.因此,采用 MEFS 方法进行空间分类可以提高分类处理的质量.

Table 2 Comparison of classification in time and quality

表 2 分类时间和质量的比较

Number of predicates	Times [*s*] (ID3)	Times [*s*] (ME)	Quality (ID3) (%)	Quality (ME) (%)
8	798	584	89.2	92.4
4	334	179	79.6	82
1	97	67	76.5	80.7

4 结 语

本文提出一种基于最大熵理论的空间特征选择方法.该方法充分考虑了对象间的相互表征作用和最大熵理论的良好统计分析特性,有效地求得了表征空间对象的特征属性.实验中利用特征分析结果进行空间分类,取得了较好的效果.我们正在深入地将这一理论应用到空间分类.

References:

[1] Oliver R, Ralf K, Simon F, Ingo M. A hybrid approach to feature selection and generation using an evolutionary algorithm. Technical Report, CI-127/02, Collaborative Research Center 531,University of Dortmund, 2002.

[2] Jain AK, Duin RPW, Mao J. Statistical pattern recognition: A review. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000,22(1):4~37.

[3] Yao X. Evolving artificial neural networks. Proceedings of the IEEE, 1999,87(9):1423~1447.

[4] Raman B, Ioerger TR. Instance based filter for feature selection. Journal of Machine Learning Research, 2002,(1):1~23. http://citeseer.nj.nec.com/raman02instance.html.

[5] Kohavi R, John G. Wrappers for feature subset selection. Artificial Intelligence, 1997,97(1-2).

[6] Koperski K, Han J, Stefanovic N. An efficient two-step method for classification of spatial data. In: Proceedings of the International Symposium on Spatial Data Handling (SDH'98). Vancouver, 1998. 45~54.

[7] Smadja F. Retrieving collocation from text: Xtract. Computational Linguistics, 1993,19(1):143~175.

[8] Church KP. Word association norms, mutual information, and lexicography. Computational Linguistics, 1990,16(1):22~29.

[9] Berger A. The improved iterative scaling algorithm: A gentle introduction. 1997. http://citeseer.nj.nec.com/31826.html.