数据库与信息处理

一种基于信息增益的K-NN改进算法

魏孝章[1]，豆增发[2]

1.陕西教育学院 数理工程系，西安 710061
2.西安电子科技大学 计算机学院，西安 710071

摘要　　针对传统K-NN算法易受单个属性干扰和时间效率较低的问题，提出了利用信息增益和可拓关联度对其进行改进。通过计算属性的信息增益来确定属性的权重系数，根据权重系数将属性划分为关键属性、次要属性和无关属性，在计算欧氏距离时引入权重系数，使各个属性的作用受其重要性的约束，有效地提高了K-NN算法的抗干扰能力和精确性。将属性空间划分为若干个子空间，利用可拓关联度将待测样本映射到某个子空间中，由这个子空间组成搜索空间，减少计算量，提高时间效率;测试结果表明，改进后的算法可行有效。

关键词　　K-NN算法　 信息增益　 信息熵　 可拓关联度

分类号

Enhancement of K-Nearest Neighbour algorithm using information gain

WEI Xiao-zhang[1]，DOU Zeng-fa[2]

1.Department of Mathematical Engineering，Education College of Shaanxi，Xi'an 710061，China
2.School of Computer Science，Xidian University，Xi'an 710071，China

Abstract

　Because the conventional K-NN algorithm has the disadvantages that the result is easily disturbed by single attribute of record and the time efficiency is very low，a novel K-NN algorithm is proposed which is an enhanced K-NN algorithm using information gain and extension relativity.Get the role coefficient of attribute，by which the attributes can be classified into primary attribute，secondary attribute and redundant attribute，through computing the information gain of attribute.The role coefficient is introduced to the Euler's distance equation so that the effect of every attribute is controled by weightiness of it，by that，the anti-jamming ability and accuracy of the K-NN algorithm are improved highly.Partition the attribute space into several sub-spaces，and map the record，which will be recognized，to a sub-space which a searching space is comprised of.So，the computing time is decreased and the time efficiency is improved greatly.The test results show the novel K-NN algorithm is feasible and effective.

Key words　K-Nearest Neighbour algorithm　 information gain　 information entropy　 extension relativity

DOI:

通讯作者　魏孝章 E-mail：jssdzf@126.com

扩展功能

本文信息
- Supporting info
- PDF(1012KB)
- [HTML全文](0KB)
- 参考文献

服务与反馈
- 把本文推荐给朋友
- 加入我的书架
- 加入引用管理器
- 复制索引
- Email Alert
- 文章反馈
- 浏览反馈信息

相关信息
- 本刊中 包含"K-NN算法"的相关文章
- 本文作者相关文章
- 魏孝章
- 豆增发