海量数据干扰下的危险Web数据挖掘技术研究

王曙霞，熊曾刚
（湖北工程学院 计算机与信息科学学院，湖北 孝感 432000）

摘 要：提出一种海量数据干扰下基于自组织映射的危险web数据挖掘算法，通过海量数据的预测值与实际值之间的误差对其中的干扰数据进行判断和排除，在此基础上，通过自组织特征映射网络对危险web数据进行挖掘，介绍了自组织特征映射网络及输出层竞争的详细过程，确定一个可形成映射的网络，将待挖掘危险web数据看作是输入向量输入自组织映射网络中，在输出图上产生相应的胜出点，将相似的输入向量汇聚在映射图的相邻区域，与该区域距离较远的胜出点对应的输入向量则可被判断是危险web数据. 仿真实验结果表明，采用所提算法对海量数据干扰下的危险web数据进行挖掘，不仅具有很高的挖掘效率，而且在挖掘精度上也有很高的性能.

关键词：海量数据；干扰；危险web数据；挖掘
中图分类号：TP311 **文献标识码**：A **文章编号**：1000－7180(2016)02－0087－05

Huge Amounts of Data Under the Interference of Dangerous Web Data Mining Technology Research

WANG Shu-xia, XIONG Zeng-gang

(School of Computer and Information Science, Hubei Engineering University, Xiaogan 432000, China)

Abstract: Put forward a lot of data under the dangerous web data mining algorithm based on self-organizing mapping, through massive amounts of data error between the predicted values and the actual value of the judgment and exclude the interference of data, on this basis, through self-organization feature mapping networks for dangerous web data mining, self-organizing feature map network was introduced and the detailed process of the output layer competition, determine a network, can form mapping will be dangerous web data mining as a self-organizing map network input vector input, the output wins produced the corresponding points on the graph, similar input vector convergence in adjacent areas of the map, distance and the region's victory points corresponding to the input vector is dangerous web data can be judge. The simulation results show that the proposed algorithm for huge amounts of data under the interference of dangerous web data mining, not only has high efficiency, and also has high performance in mining precision.

Key words: huge amounts of data; interference; dangerous web data; mining

作者简介：

王曙霞 女，(1975-)，硕士，副教授. 研究方向为智能计算与网络安全. E-mail: wsxwsxxia@163.com

熊曾刚 男，(1974-)，博士，教授. 研究方向为对等计算、网格计算、云计算和信息系统分析与集成.