

# Author's Accepted Manuscript

Feature selection based on quality of information

Jinghua Liu, Yaojin Lin, Menglei Lin, Shunxiang Wu, Jia Zhang



PII: S0925-2312(16)31299-1  
DOI: <http://dx.doi.org/10.1016/j.neucom.2016.11.001>  
Reference: NEUCOM17694

To appear in: *Neurocomputing*

Received date: 15 January 2015  
Revised date: 8 June 2016  
Accepted date: 1 November 2016

Cite this article as: Jinghua Liu, Yaojin Lin, Menglei Lin, Shunxiang Wu and Jia Zhang, Feature selection based on quality of information, *Neurocomputing* <http://dx.doi.org/10.1016/j.neucom.2016.11.001>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

## Feature selection based on quality of information

Jinghua Liu<sup>a,b</sup>, Yaojin Lin<sup>b,\*</sup>, Menglei Lin<sup>b</sup>, Shunxiang Wu<sup>a</sup>, Jia Zhang<sup>b</sup>

<sup>a</sup>Department of Automation,  
Xiamen University, Xiamen, 361000, P. R. China

<sup>b</sup>School of Computer Science,  
Minnan Normal University, Zhangzhou, 363000, P. R. China

---

**Abstract**

Feature selection as one of the key problems of data preprocessing is a hot research topic in pattern recognition, machine learning, and data mining. Evaluating the relevance between features based on information theory is a popular and effective method. However, very little research pays attention to the distinguishing ability of feature, i.e., the degree of a feature distinguishes a given sample with other samples. In this paper, we propose a new feature selection method based on the distinguishing ability of feature. First, we define the concept of maximum-nearest-neighbor, and use this concept to discriminate the nearest neighbors of samples. Then, we present a new measure method for evaluating the quality of feature. Finally, the proposed algorithm is tested on benchmark datasets. Experimental results show that the proposed algorithm can effectively select a discriminative feature subset, and performs as well as or better than other popular feature selection algorithms.

*Keywords:* Feature selection, Information entropy, Maximum-nearest-neighbor, Quality of information

---

**1. Introduction**

Feature selection as an important data preprocessing technique is widely applied in pattern recognition and machine learning. With the development of information storage, input data have a large number of features, which may include a mass of irrelevant and/or redundant features. Unnecessary features often result in "dimensionality curse", the low efficiency of learning algorithm, and the problem of over-fitting [5, 13, 25, 40]. Therefore, it is an extremely important step to select the relevant and necessary features for a given learning task.

Feature selection is a process of selecting an optimal feature subset from the raw feature

---

\*Corresponding author.

*Email addresses:* zzliujinghua@163.com (Jinghua Liu), yjlin@mnnu.edu.cn (Yaojin Lin), menglei36@126.com (Menglei Lin), sxwu@xmu.edu.cn (Shunxiang Wu), zhangjia\_g1@163.com (Jia Zhang)

Download English Version:

<https://daneshyari.com/en/article/4948059>

Download Persian Version:

<https://daneshyari.com/article/4948059>

[Daneshyari.com](https://daneshyari.com)