

## Accepted Manuscript

Nearest neighbor classification of categorical data by attributes weighting

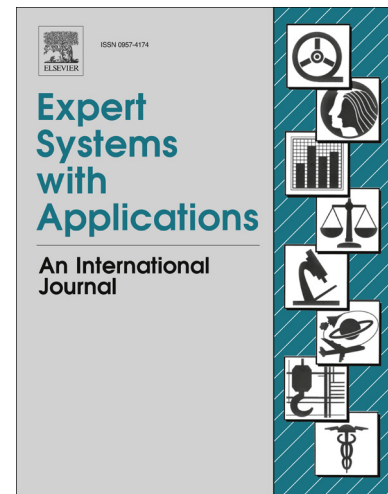
Lifei Chen, Gongde Guo

PII: S0957-4174(14)00767-2

DOI: <http://dx.doi.org/10.1016/j.eswa.2014.12.002>

Reference: ESWA 9723

To appear in: *Expert Systems with Applications*



Please cite this article as: Chen, L., Guo, G., Nearest neighbor classification of categorical data by attributes weighting, *Expert Systems with Applications* (2014), doi: <http://dx.doi.org/10.1016/j.eswa.2014.12.002>

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 proof before it is published in its final 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.

# Nearest Neighbor Classification of Categorical Data by Attributes Weighting

Lifei Chen<sup>a,b</sup>, Gongde Guo<sup>a</sup>

<sup>a</sup>*School of Mathematics and Computer Science, Fujian Normal University, China*

<sup>b</sup>*Department of Computer Science, University of Sherbrooke, Canada*

---

## Abstract

Subspace classification of categorical data is an essential process for many real-world applications such as computer-aided medical diagnosis and collaborative recommendation. The nearest neighbor classifiers have sparked wide interest from these applications because of their simplicity and flexibility. However, they become ineffective when applied to categorical data, due to the lack of a well-defined distance measure used to compute dissimilarities between categorical samples in the projected subspaces. In this paper, we tackle the problem by defining a series of weighted distance functions for categorical attributes, and applying them to derive new nearest neighbor classifiers. Four attribute-weighting measures are proposed, with two defined on global feature-ranking approaches while the other two on local approaches. The experimental results conducted on real categorical data sets demonstrate that all four classifiers outperform consistently the traditional methods, and show the suitability of the proposal for the real applications in terms of automated feature selection.

**Keywords:** Nearest neighbor classification, categorical data, distance measure, projected subspace, feature selection, attribute weighting

---



---

*Email addresses:* `ggd@fjnu.edu.cn` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo), `lifei.chen@usherbrooke.ca` (Gongde Guo), `ggd@fjnu.edu.cn` (Gongde Guo)

Download English Version:

<https://daneshyari.com/en/article/10321867>

Download Persian Version:

<https://daneshyari.com/article/10321867>

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