

Accepted Manuscript

Interaction-based feature selection using Factorial Design

Xiaochuan Tang, Yuanshun Dai, Peng Sun, Sa Meng

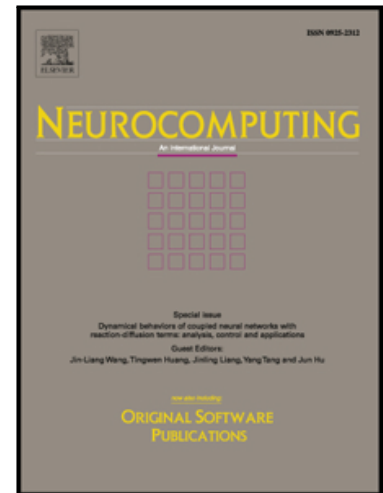
PII: S0925-2312(17)31817-9
DOI: [10.1016/j.neucom.2017.11.058](https://doi.org/10.1016/j.neucom.2017.11.058)
Reference: NEUCOM 19123

To appear in: *Neurocomputing*

Received date: 29 September 2016
Revised date: 26 September 2017
Accepted date: 25 November 2017

Please cite this article as: Xiaochuan Tang, Yuanshun Dai, Peng Sun, Sa Meng, Interaction-based feature selection using Factorial Design, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.11.058](https://doi.org/10.1016/j.neucom.2017.11.058)

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.



Interaction-based feature selection using Factorial Design

Xiaochuan Tang, Yuanshun Dai*, Peng Sun, Sa Meng

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Abstract

Feature interaction provides insight into hidden domain knowledge and interactive structure of a data set. In feature selection, identifying significant interactions among features is a challenging task. Since possible candidates of interactions increase exponentially to the number of features. In this paper, we propose a two-stage feature selection approach that makes full use of interactions. In the first stage, we decompose the feature selection problem into a sum of interaction information. Then, higher-order interactions are used to select an interaction-preserving feature subset. In the second stage, we employ design of experiments (DOE) to identify significant interactions from the feature subset. The proposed approach is compared with mRMR, JMIM and ReliefF. Experiments on public available data sets demonstrate that our approach reveals the influence of interactions and so that outperforms the state-of-the-art filter methods.

Keywords: Feature selection, Interaction, Mutual information, Factorial design, Design of experiments

1. Introduction

Vast amount of data is being generated in many fields, such as text mining, image processing and bioinformatics [1]. However, techniques for analyzing higher dimensional data suffer from the multidimensional complexity. Toward this issue, a variety of data preprocessing methods have been proposed. These methods can be mainly divided into two categories: feature selection (FS) and dimensionality reduction. Feature selection identifies a characteristic feature subset. Thus, it is more explicable than dimensionality reduction. Feature selection techniques can be grouped into three categories: filter[2], wrapper[3] and learning based[4, 5, 6]. Many learning based methods has been proposed,

*Corresponding author

Email addresses: tangxc.uestc@gmail.com (Xiaochuan Tang), ydai@uestc.edu.cn (Yuanshun Dai), sleisurep@163.com (Peng Sun), summerincuit@gmail.com (Sa Meng)

Download English Version:

<https://daneshyari.com/en/article/6864622>

Download Persian Version:

<https://daneshyari.com/article/6864622>

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