Journal of IT in Asia, *Vol. 6*, *2016*, *11 – 24*

Improved Feature Selection Based on Mutual Information for Regression Tasks

[1]Muhammad A. Sulaiman and [2]Jane Labadin
[1,2]Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, 94300 Kota Samarahan, Sarawak, Malaysia
email: [1]muhalisu@gmail.com, [2]ljane@unimas.my

Abstract – *Mutual Information (MI) is an information theory concept often used in the recent time as a criterion for feature selection methods. This is due to its ability to capture both linear and non-linear dependency relationships between two variables. In theory, mutual information is formulated based on probability density functions (pdfs) or entropies of the two variables. In most machine learning applications, mutual information estimation is formulated for classification problems (that is data with labeled output). This study investigates the use of mutual information estimation as a feature selection criterion for regression tasks and introduces enhancement in selecting optimal feature subset based on previous works. Specifically, while focusing on regression tasks, it builds on the previous work in which a scientifically sound stopping criteria for feature selection greedy algorithms was proposed. Four real-world regression datasets were used in this study, three of the datasets are public obtained from UCI machine learning repository and the remaining one is a private well log dataset. Two Machine learning models namely multiple regression and artificial neural networks (ANN) were used to test the performance of IFSMIR. The results obtained has proved the effectiveness of the proposed method.*

Keywords: feature selection, filter, estimation, mutual information, machine learning.

1 Introduction

Feature Selection is an optimization method used to select optimal subsets of a full feature set. The selected optimal subset is expected to retain the relevant information in the full feature set. Feature selection algorithms rely on some certain criteria to score features or input variables, these criteria are divided into two categories and thereby forming the basis for the three types of feature selection models. The first category of criteria utilizes statistical and a probabilistic distribution of dataset attributes to measure the relevant of each input variable to the corresponding output variable. This category is refers to as filter models. The second category of criteria is search optimization algorithms used together with a particular learning machine model to find relevant feature subset based on the performance of the learning machine. This category comprises of wrapper models. And the third type of feature selection model is the hybrid models, which combined the power of both filter and wrapper to select feature subsets.

Mutual information is used as a criterion for feature selection method. It is a filter-based model that measures both linear and non-linear dependency relationships between two random variables, this property made it be a popular choice as feature selection criterion. Mutual information is formulated from probability density functions (pdfs) or entropies of two variables and their joint variable. However, despite that in theory mutual information formulation is suitable for a dataset with either discrete or continuous output variable, in practice its estimation is often assumed classification problems (that is a dataset with labeled output variable). The reason for this cannot be unconnected to the fact that it is not clear how or rather very difficult to estimate pdf or entropy of joint variable from a dataset with a continuous output variable. This study presents mutual information estimation formulated around the estimated entropy of a variable. In addition, this study provides an extension to feature selection greedy algorithms on how to select optimal feature subset when using filter model only.

The rest of the paper is organized as follows: Section 2 is a review of feature selection methods, information theory and how it relates to mutual information. Section 3 presents Machine learning models used in this study. Feature selection procedure based on MI estimation is presented in section 4. Section 5 presents experimental studies, results and discussion. And finally, the conclusion is presented in section 6.