Fuzzy Mutual Information Feature Selection Based on Representative Samples

Omar A. M. Salem, Faculty of Computer Science and Informatics, Suez Canal University, Ismailia, Egypt

Liwei Wang, International School of Software, Wuhan University, Wuhan, China

ABSTRACT

Building classification models from real-world datasets became a difficult task, especially in datasets with high dimensional features. Unfortunately, these datasets may include irrelevant or redundant features which have a negative effect on the classification performance. Selecting the significant features and eliminating undesirable features can improve the classification models. Fuzzy mutual information is widely used feature selection to find the best feature subset before classification process. However, it requires more computation and storage space. To overcome these limitations, this paper proposes an improved fuzzy mutual information feature selection based on representative samples. Based on benchmark datasets, the experiments show that the proposed method achieved better results in the terms of classification accuracy, selected feature subset size, storage, and stability.

KEYWORDS

Feature Selection, Fuzzy Mutual Information, Fuzzy Sets, Mutual Information

INTRODUCTION

Nowadays, classification models have various applications in many areas such as medical, business, engineering, life and social sciences. As the size of real-world datasets from these areas continues to increase, building classification models become a significantly more difficult task (Janecek et al., 2008). Although high-dimensional data include important features, it may also include undesirable data such as irrelevant and redundant features. The presence of undesirable features leads to a decrease in classification accuracy (Dash and Liu, 2003; Vieira et al., 2012). Moreover, it increases storage space and memory usage (Dash and Liu, 2003; Janecek et al., 2008). So, selecting relevant features and eliminating irrelevant or redundant features helps to build effective classification models (Yu et al., 2011).

Features selection as a preprocessing step aims to select the minimum subset that describes the data efficiently and increases the classification accuracy (Guyon and Elisseeff, 2003). It can be grouped into a wrapper, filter, and embedded approaches. Both wrapper and embedded approaches can be considered as classifier-dependent feature selection, while filter approaches can be considered as a classifier-independent feature selection (Bennasar et al., 2015). In this study, we use filter approach according to its advantages over wrapper or embedded approaches. The main advantages of filter

DOI: 10.4018/IJSI.2018010105

approaches are classifier-independent, less time consuming and more practical for classification models (Saeys et al., 2007).

Filter approaches try to filter undesirable features out before classification process (Garc´ıa et al., 2015). They select the highly ranked features based on characteristics of the training data (Guyon and Elisseeff, 2003). The main characteristics of data depend on two relations: relevance and redundancy (Chandrashekar and Sahin, 2014). Relevance describes how the features can discriminate the different classes, while redundancy describes how the features depend on each other. So maximizing feature relevance and minimizing feature redundancy leads to best feature ranking. To evaluate the characteristics of features, filter approach uses many evaluation measures such as correlation (Hall, 1999), Shannon mutual information (Vergara and Est´evez, 2014). Correlation measures are suitable only for a linear relationship among features, while Shannon mutual information is suitable for linear and non-linear relations among features (Lee et al., 2012). However, Shannon mutual information has some limitations: First, it requires discretization step before dealing with continuous data. But, it is difficult to avoid information loss results from discretization (Ching et al., 1995; Shen and Jensen, 2004). Second, it depends only on the inner-class information without considering outer-class information (Liang et al., 2002).

To overcome these limitations, various algorithms based on mutual information with fuzzification has been introduced in many literatures. Yu et al. (2011). proposed a fuzzy mutual information using logarithmic concept. Another algorithm was proposed to estimate a fuzzy mutual information using complement instead of logarithmic concept (Zhao et al., 2015). Both of fuzzy mutual information algorithms depend on the fuzzy binary relation. This relation can be represented in relation matrix. The size of relation matrix depends on the number of samples in the input feature. Each row or column in the relation matrix represents the relation between one sample and each of the remaining samples. So, estimating relation matrix requires more storage and computational time, especially for datasets with a tremendous amount of samples (Yu et al., 2007). Motivated by these limitations of fuzzy mutual information, we proposed a new estimation of relation matrix. To create this matrix, we estimated the relation between one sample and representative samples. These samples consist of the averages of data samples belonging to the same class. Using representative samples instead of all samples can reduce the size of relation matrix.

This paper is organized as follows: In the following section, we proposed fuzzy mutual information feature selection. Next, the feature selection evaluation measures are defined. Then we set up the experiment. After that, the experimental results are discussed. Finally, the paper is concluded.

FUZZY MUTUAL INFORMATION

Information measures, such as entropy and mutual information, are widely used in filter approaches (Battiti, 1994). These measures can be extended to define fuzzy entropy and fuzzy mutual information. Fuzzy entropy is the average amount of uncertainty which measures the discriminative power of fuzzy relations (Hu et al., 2007, 2006), while fuzzy mutual information measures the relevance and redundancy among fuzzy relations (Battiti, 1994; Hu et al., 2007). In the following, we propose a new estimation of relation matrix for fuzzy mutual information based on representative samples using logarithmic and complement concepts. Then mRMR approach is presented. Lastly, we integrate normalized fuzzy mutual information with mRMR.

Fuzzy Mutual Information Based on Representative Samples

Given a finite set of n elements $X = \{x_1, x_2, ..., x_n\}$, A clustering C of X divides it into k set of classes $C_1, ..., C_k$, where $X = \bigcup_{j=1}^{k} C_j$. Let the average of all samples belonging to class C_j is represented by v_j. The similarity measure between x_i and v_j is defined by a fuzzy relation $\bar{R}$ on X: