Expert Systems with Applications 36 (2009) 4517–4522

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Fuzzy classification systems based on fuzzy information gain measures

Shyi-Ming Chen [a,b,*], Jen-Da Shie [a]

[a] Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC
[b] Department of Computer Science and Information Engineering, Jinwen University of Science and Technology, Taipei County, Taiwan, ROC

ARTICLE INFO

Keywords:
Fuzzy information gain
Fuzzy entropy
Classification problems
Feature weights
Membership grades

ABSTRACT

In this paper, we present a new method for handling classification problems using a new fuzzy information gain measure. Based on the proposed fuzzy information gain measure, we propose an algorithm for constructing membership functions, calculating the class degree of each subset of training instances with respect to each class and calculating the fuzzy entropy of each subset of training instances. Based on the constructed membership function of each fuzzy set of each feature, the obtained class degree of each subset of training instances with respect to each class and the obtained fuzzy entropy of each subset of training instances, we propose an evaluating function for classifying testing instances. The proposed method gets higher average classification accuracy rates than the methods presented in [John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In *Proceedings of the 11th conference on uncertainty in artificial intelligence, Montreal, Canada* (pp. 338–345); Platt, J. C. (1999). Using analytic QP and sparseness to speed training of support vector machines. In *Proceedings of the 13th annual conference on neural information processing systems, Denver, Colorado* (pp. 557–563); Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Francisco: Morgan Kaufmann].

1. Introduction

Classification techniques have been widely applied in many domains. Many kinds of classifiers have been proposed for dealing with classification problems, such as the rule-based method (Banerji, 1964; Chen & Chang, 2005; Chen, Kao, & Yu, 2002; Gomez, Garcia, & Silva, 2005), the instance-based method (Cover & Hart, 1967), the linear function method (Fisher, 1936), the decision trees method (Quinlan, 1986), the C4.5 method (Quinlan, 1993), the artificial neural networks method (ANN) (McCulloch & Pitts, 1943), the support vector machines (SVM) method (Boser, Guyon, & Vapnik, 1992), the sequential minimal optimization (SMO) method (Platt, 1999), the naive Bayes method (John & Langley, 1995), the genetic algorithm method (Chen & Chen, 2002; Winkler, Affenzeller, & Wagner, 2006), etc. A data set might have numeric or nominal features. Some classifiers may have high classification accuracy rates for numeric data and some other classifiers may have high classification accuracy rates for nominal data. If a classifier can only deal with nominal features, the discretization of numeric features is necessary. If we can properly discretize each numeric feature into finite subsets, then we can increase the classification accuracy rates of classification systems (Catlett, 1991).

In this paper, we propose a new method for handling classification problems based on a new fuzzy information gain measure. First, we propose a new fuzzy information gain measure for a feature with respect to a set of training instances. Then, based on the proposed fuzzy information gain measure, we propose an algorithm for constructing membership functions, calculating the class degree of each subset of training instances with respect to each class and calculating the fuzzy entropy of each subset of training instances, where each subset of training instances contains a part of the training instances whose values of a specific feature fall in the support of a specific fuzzy set of this feature. Finally, based on the constructed membership function of each fuzzy set of each feature, the obtained class degree of each subset of training instances with respect to each class and the obtained fuzzy entropy of each subset of training instance, we propose an evaluating function for classifying testing instances. The proposed gets higher average classification accuracy rates than the ones presented in John and Langley (1995), Platt (1999), and Quinlan (1993).

The rest of this paper is organized as follows. In Section 2, we present a method to construct membership functions of a numeric feature. In Section 3, we briefly review the existing entropy measures and propose a new fuzzy information gain measure for a feature with respect to a set of training instances. In Section 4, we propose a new method for dealing with classification problems based on the proposed fuzzy information gain measure. In Section 5, we compare the average classification accuracy rates of the proposed method with the ones using the existing methods (John &

* Corresponding author. Address: Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, ROC. Tel.: +886 2 27376417; fax: +886 2 27301081.
E-mail address: smchen@mail.ntust.edu.tw (S.-M. Chen).

0957-4174/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2008.05.020

Langley, 1995; Platt, 1999; Quinlan, 1993). The conclusions are discussed in Section 6.

2. Construct membership functions of each numeric feature

In this section, we propose a method for constructing membership functions of a numeric feature. Assume that R denotes a set of training instances and f denotes a feature. First, we apply the k-means clustering algorithm (Hartigan & Wong, 1979) for clustering the values of the numeric feature f of a set of training instances into k clusters (i.e., $[x_{11},x_{12}]$, $[x_{21},x_{22}]$, …, $[x_{k1},x_{k2}]$) and generate k cluster centers $m_1, m_2, \ldots, m_k$. Then, we use these cluster centers $m_1, m_2, \ldots, m_k$ as the centers of the fuzzy sets to construct the membership functions of the fuzzy sets of the feature f as follows:

for i = 1 **to** k **do**
{
if i = 1 **then**
{
/* Construct the membership function μ_{v_1} corresponding to the first cluster center m_1, where $U_{\min}$ denotes the minimum value of the universe of discourse of the feature f, m_1 denotes the first cluster center, and m_2 denotes the second cluster center. */

let $$\mu_{v_1}(x) = \begin{cases} 1 - \frac{x-m_1}{U_{\min}-m_1} \times 0.5, & \text{if } U_{\min} \leqslant x \leqslant m_1, \\ 1 - \frac{x-m_1}{m_2-m_1}, & \text{if } m_1 < x \leqslant m_2, \\ 0, & \text{otherwise}, \end{cases}$$

}
else
{
if i = k **then**
{
/* Construct the membership function μ_{v_k} corresponding to the kth cluster center m_k, where $U_{\max}$ denotes the maximum value of the universe of discourse of the feature f, m_k denotes the kth cluster center, m_{k-1} denotes the left cluster center of m_k. */

let $$\mu_{v_k}(x) = \begin{cases} 1 - \frac{x-m_k}{m_{k-1}-m_k}, & \text{if } m_{k-1} \leqslant x \leqslant m_k, \\ 1 - \frac{x-m_k}{U_{\max}-m_k} \times 0.5, & \text{if } m_k < x \leqslant U_{\max}, \\ 0, & \text{otherwise}, \end{cases}$$

}
else
{
/* Construct the membership function μ_{v_i} corresponding to the ith cluster center m_i, where $1 < i < k$, m_i denotes the ith cluster center, m_{i-1} denotes the left cluster center of m_i, and m_{i+1} denotes the right cluster center of m_i. */

let $$\mu_{v_i}(x) = \begin{cases} 1 - \frac{x-m_i}{m_{i-1}-m_i}, & \text{if } m_{i-1} \leqslant x \leqslant m_i, \\ 1 - \frac{x-m_i}{m_{i+1}-m_i}, & \text{if } m_i < x \leqslant m_{i+1}, \\ 0, & \text{otherwise}. \end{cases}$$

}
}
}.

3. Fuzzy information gain measures

In this section, we propose a fuzzy information gain measure of each feature with respect to a set of training instances. In the following, we propose the definition of the fuzzy information gain measure of a feature with respect to a set of training instances. A feature can be described by several linguistic terms (Zadeh, 1975) and each linguistic term can be represented by a fuzzy set (Zadeh, 1965) characterized by a membership function. Assume that a set R of training instances is classified into a set C of classes and assume that a set V of fuzzy sets is defined in the feature f. Let μ_v denote the membership function of the fuzzy set v, $v \in V$, and let $\mu_v(x)$ denote the membership grade of value x belonging to the fuzzy set v. The support U_v of a fuzzy set v denotes a subset of the universe of discourse U of a feature, where $U_v = \{u | \mu_v(u) > 0 \text{ and } u \in U\}$. First, we review a class degree measure for calculating the degree of possibility of a subset of training instances belonging to a specific class whose values of a specific feature fall in the support of a specific fuzzy set, shown as follows.

Definition 3.1. (Chen & Shie, 2005) The class degree $CD_c(v)$ of a subset R_v of training instances belonging to a class c whose values of a feature f fall in the support U_v of a fuzzy set v of the feature f is defined by

$$CD_c(v) = \frac{\sum_{x \in X_c} \mu_v(x)}{\sum_{x \in X} \mu_v(x)}, \tag{1}$$

where X denotes the set of values of the feature f of the subset R_v of training instances, $X \subset U_v$, X_c denotes the set of values of the feature f of the subset R_v of training instances belonging to the class c, $c \in C$, μ_v denotes the membership function of the fuzzy set v, $\mu_v(x)$ denotes the membership grade of value x belonging to the fuzzy set v, and $\mu_v(x) \in [0,1]$.

Then, based on the class degree measure shown in Eq. (1), we review a fuzzy entropy measure for calculating the uncertainty about the class of a subset of training instances whose values of a specific feature fall in the support of a specific fuzzy set, shown as follows.

Definition 3.2. (Chen & Shie, 2005; Shie & Chen, 2007) The fuzzy entropy $FE(v)$ of a subset of training instances whose values of a feature f fall in the support U_v of a fuzzy set v of the feature f is defined by

$$FE(v) = -\sum_{c \in C} CD_c(v) \log_2 CD_c(v), \tag{2}$$

where C denotes a set of classes and $CD_c(v)$ denotes the class degree of the subset of training instances belonging to a class c whose values of the feature f fall in the support U_v of the fuzzy set v of the feature f.

Then, based on the fuzzy entropy measure shown in Eq. (2), we propose a fuzzy information gain measure for calculating the expected reduction in entropy caused by partitioning the set of training instances according to a specific feature, shown as follows.

Definition 3.3. (Shie & Chen, 2006) The fuzzy information gain $FIG(R,f)$ of a feature f with respect to a set R of training instances is defined by

$$FIG(R,f) = \sum_{c \in C} \left(-\frac{n_c}{n} \log_2 \frac{n_c}{n} \right) - \sum_{v \in V} \left(\frac{S_v}{S} FE(v) \right), \tag{3}$$

where n denotes the number of instances contained in R, n_c denotes the number of instances which are contained in R and belonging to a class c, V denotes the fuzzy sets of the feature f, s denotes the summation of the membership grades of the values of the feature f of the set R of training instances belonging to each fuzzy set of the feature f, S_v denotes the summation of the membership grades of the values of the feature f of the set R of training instances belonging to a fuzzy set v of the feature f, and $FE(v)$ denotes the fuzzy entropy of a subset of training instances whose values of the feature f fall in the support U_v of a fuzzy set v of the feature f.