Mutual Information in Naive Bayes with Kernel Density Estimation

Zhongliang Xiang 1,2, Xiangru Yu 1, and Dae-Ki Kang 2,*

1Weifang University of Science & Technology, 2Dongseo University

E-mail : dkkang@dongseo.ac.kr *

ABSTRACT

Naive Bayes (NB) assumption has some harmful effects in classification to the real world data. To relax this assumption, we now propose approach called Naive Bayes Mutual Information Attribute Weighting with Smooth Kernel Density Estimation (NBMIKDE) that combine the smooth kernel for attribute and attribute weighting method based on mutual information measure.

Keywords

Naive Bayes, Attribute Weighting, Mutual Information, Kernel Density Estimation.

Ⅰ. Introduction

The NB classifier is a certain supervised learning method based on Bayes’ rule on statistic theory scale, running on labeled training examples and given by a strong assumption that all attributes in training example are independent to each other on given the training examples, so-called Naive Bayes assumption or Naive Bayes conditional independent assumption. NB classifier has high performance and rapidly classifies speed, its effectiveness especially in a huge of training instances with plenty of attributes. Those benefits come from NB assumption sometimes.

Many researchers provide proposals to relax NB assumption effectively. Jiang and Wang et al. [1] made a survey about improving NB methods, those improving methods are broadly divided into five main categories: structure extension, feature selection, data expansion, local learning, and attribute weighting. In this paper, we focus on attribute weighting way and combine Kernel Density Estimation (KDE) in NB learning to relax conditional independent assumption.

Ⅱ. Kernel Density Estimation for Naive Bayes Categorical Attributes

In NB learning, the posterior probability $p(a_i|c)$ is often estimated by $f_c(a_i)$, the frequency of $a_i$ given c. From a statistical perspective, a non-smooth estimator has the least sample bias, but it also has a large estimation variance [2, 3] at same time. Aitchison et al. [4] proposed a kernel function, and Lifei Chen et al. [3] also proposed a variant smooth kernel function for frequency from [4].

Ⅲ. Our Proposed Approach

In this section, we propose our novel
approach named Naive Bayes Mutual Information Attribute Weighting with Smooth Kernel Density Estimation (NBMIKDE). NBMIKDE made a combination which employing mutual information to attribute weighting and kernel density estimation for categorical attribute to relax NB assumption. In the part of attributes weighting in our approach, we will generate a series $w_i$ for each $A_i$, the details in 3.2 section be shown.

### 3.1 NBMIKDE Kernel Density Estimator

From Lifei Chen et al. [3] it make sense that when given a class, if someone attribute $A_i$ has more important for classify, in other word, $A_i$ can provide more classify ability than other attributes, so this attribute should be assigned more large weight. The average weight $\bar{w}_i$ of each attribute $A_i$ is defined as follows:

$$\bar{w}_i = \frac{\sum_{j=1}^{m} I(A_i; c)}{\sum_{j=1}^{m} I(A_i; c)}$$  \hspace{1cm} (4)

where the definition of $I$ is follows:

$$I(A_i; c) = \sum_{j} p(a_j | c) \log \frac{p(a_j | c)}{p(a_j)}$$  \hspace{1cm} (5)

We also imply split information same like C4.5 [8] to avoid choosing the attributes with lots of values. The split information for each $A_i$ is defined as follows:

$$\text{Split}(A_i) = -\sum_{j} p(a_j | c) \log p(a_j | c)$$  \hspace{1cm} (6)

Now, the weight of $A_i$ is defined as follows:

$$w_i = \frac{\text{Split}(A_i)}{\sum_{i=1}^{n} \text{Split}(A_i)}$$  \hspace{1cm} (7)

The Output for NBMIKDE is:

$$c(t) = \arg \max_{c \in C} \left( p(c) \prod_{i=1}^{m} p(a_i | c, w_i) \right)$$

### 4. Conclusion

In this paper, we proposed a novel approach: Naive Bayes Mutual Information Attribute Weighting with Smooth Kernel Density Estimation. NBMIKDE made a combination which employing mutual information to attribute weighting and kernel density estimation for categorical attribute to relax NB assumption.

### References


