나이브 베이스에서의 커널 밀도 측정과 상호 정보량

샹총량[1,2], 유샹루[1], 강대기[2,*]

[1]웨이팡과기대학, [2]동서대학교

Mutual Information in Naive Bayes with Kernel Density Estimation

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

[1]Weifang University of Science & Technology, [2]Dongseo University

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

요 약

나이브 베이스가 가지는 가정은 실세계 데이터를 분류함에 있어 해로운 효과를 보이곤 한다. 이러한 가정을 완화하기 위해, 우리는 Naive Bayes Mutual Information Attribute Weighting with Smooth Kernel Density Estimation (NBMIKDE) 접근 방법을 소개한다. NBMIKDE는 애트리뷰트를 위한 스무드 커널과 상호 정보량 측정값을 기반으로 하는 어트리뷰트 가중치 기법을 조합한 것이다.

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.

키워드

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 $\overline{f_c}(a_i)$, 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 provides more information to reduce the indeterminacy of class c, then the value of $p(a_i|c)$ should more close to $\overline{f_c}(a_i)$, otherwise, if A_i has less meaning for classify, then $p(a_i|c)$ should more close to $\frac{1}{|A_i|}$, and $|A_i|$ is the cardinality of attribute. We let the bandwidth $\lambda_{c_i} = 1 - w_i$, the variance with [3] kernel equation as follows:

$$\kappa(t_{x_i}, a_i, w_i) = \begin{cases} \frac{1}{|A_i|} + \frac{|1-A_i|}{|A_i|} w_i, \text{if } t_{x_i} = a_i \\ \frac{1}{|A_i|}(1-w_i), \quad otherwise \end{cases} \quad (1)$$

so, the estimation $P(t_{x_i}|c,w_i)$ of probability of $P(t_{x_i}|c)$ is described as follows:

$$\begin{aligned} P(t_{x_i}|c,w_i) &= \frac{1}{n_c}\sum_{i=1}^{n_c}\kappa(t_{x_i}, a_i, w_i) \quad (2) \\ &= \frac{1}{|A_i|} + w_i\left(\overline{f_c}(t_{x_i}) - \frac{1}{|A_i|}\right) \end{aligned}$$

NBMIKDE is defined as follows:

$$c(t) = \arg max_{c \in C}\left(p(c)\prod_{i=1}^{m} p(a_i|c,w_i)\right) \quad (3)$$

3.2 NBMIKDE Attribute Weighting

We will generate the attribute weights from data. Chang-Hwan et al. [5] generated attribute weights by Kullback-Leibler Measurement. Umut Orhan et al. [6] use least squares approach to weight attributes in NB. Nayyar A. Zaidi et al. [7] proposed a weighted NB algorithm, called WANBIA, this method selects weights to minimize either the negative conditional log likelihood or the mean squared error objective functions.

Our approach generates a set of attribute weights w_i that employing mutual information between a_i and c. It makes sense that if one attribute has more mutual information value with class label, this attribute will provide more classify ability than other attributes, so this attribute should be assigned more large weight. The average weight $\overline{w_i}$ of each attribute A_i is defined as follows:

$$\overline{w_{c_i}} = \frac{I(A_i;c)}{\sum_{i=1}^{m} I(A_i;c)} \quad (4)$$

where the definition of is follows:

$$I(A_i;c) = \sum_{i,c} p(a_i|c)p(c)\log\frac{p(a_i|c)}{p(a_i)} \quad (5)$$

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:

$$Split(A_i) = -\sum_j p(a_{ij})\log(p(a_{ij})) \quad (6)$$

Now, the weight of A_i is defined as follows:

$$w_i = \frac{\frac{\overline{w_i}}{Split(A_i)}}{\sum_{i=1}^{m}\frac{\overline{w_i}}{Split(A_i)}} \quad (7)$$

The Output for NBMIKDE is:

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

IV. 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

[1] Jiang and Wang et al.: A Novel Bayes Model: Hidden Naive Bayes. IEEE Transactions on Knowledge and Data Engineering, VOL. 21, NO. 10, (2009)

[2] Qi,L., Racine, J.: Nonparametric econometrics: Theory and practice. Princeton University Press (2007)

[3] Lifei Chen, Shengrui Wang: Semi-naive Bayeian Classification by Weighted Kernel Density Estimation. (2012)
[4] Aitchison, J., Aitken, C.: Multivariate binary discrimination by the kernel method. Biometrika 63, 413-420 (1976)
[5] Chang-Hwan Lee, Fernando Gutierrez, Dejing Dou: Calculating Feature Weights in Naive Bayes with Kullback-Leibler Measure, 11th IEEE International Conference on Data Mining. (2011)
[6] Umut Orhan, Kemal Adem, Onur Comert: Least Squares Approach to Locally Weighted NB Method, Journal of New Results in Science 1,71-80 (2012)
[7] Nayyar A. Zaidi, Jesus Cerquides, Mark J. Carman, Geoffrey I. Webb: Alleviating NB Attribute Independence Assumption by Attribute Weighting, Journal of Machine Learning Research 14 1947-1988 (2013)
[8] J.Ross Quinlan. C4.5: programs for machine learning. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA. (1993)