커널 밀도 측정에서의 나이브 베이스 접근 방법

샹총량[1,2], 유샹루[1], 아메드 압둘하킴 알-압시[2], 강대기[2,*]

[1]웨이팡과기대학, [2]동서대학교

Naive Bayes Approach in Kernel Density Estimation

Zhongliang Xiang[1,2], Xiangru Yu[1], Ahmed Abdulhakim Al-Absi[2], and Dae-Ki Kang[2,*]

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

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

요 약

나이브 베이스 학습은 유명하면서도, 빠르면서도 효과적인 지도 학습 방법으로, 다소 잡음을 가진 라벨이 있는 데이터집합을 다루는 데 좋은 성능을 보인다. 그러나, 나이브 베이스의 조건적 독립성 가정은 실세계 데이터를 다루는 데 필요한 특성에 다소 제약사항을 가지게 한다. 지금까지 연구자들이 이 조건적 독립성 가정을 완화시키는 방법들을 제안해 왔다. 이러한 방법들은 어트리뷰트 가중치, 커널 밀도 측정 등이 있다. 본 논문에서, 우리는 커널 밀도 측정과 어트리뷰트 가중치를 이용하여 나이브 베이스의 학습 효과를 개선하기 위한 NB Based on Attribute Weighting in Kernel Density Estimation (NBAWKDE) 이라는 새로운 접근 방법을 제안한다.

ABSTRACT

Naive Bayes (NB, for shortly) learning is more popular, faster and effective supervised learning method to handle the labeled datasets especially in which have some noises, NB learning also has well performance. However, the conditional independent assumption of NB learning imposes some restriction on the property of handling data of real world. Some researchers proposed lots of methods to relax NB assumption, those methods also include attribute weighting, kernel density estimating. In this paper, we propose a novel approach called NB Based on Attribute Weighting in Kernel Density Estimation (NBAWKDE) to improve the NB learning classification ability via combining kernel density estimation and attribute weighting.

키워드

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

Ⅰ. Introduction

In Naive Bayes classifier, NB assumption conflicts with the fact obviously, no one can claim that attributes in same dataset must have no any relationships with each other. Therefore, 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.

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, Shengrui Wang [3] also proposed a variant smooth kernel function for frequency from [4] where with $\lambda_{c_i} \in [0,1]$ being the bandwidth. $\kappa(t_{x_i}, a_i, \lambda_{c_i})$ is a kernel function for A_i given c. This kernel function may become an indicator if $\lambda_{c_i} = 0$.

Also [3] used Equation (1) to estimate

$P(t_{x_i}|c)$ as follows:

$$P(t_{x_i}|c,\lambda_{c_i}) = \frac{1}{n_c}\sum_{i=1}^{n_c}\kappa(t_{x_i},a_i,\lambda_{c_i}) = \overline{f}_c(t_{x_i}) + \left(\frac{1}{|A_i|} - \overline{f}_c(t_{x_i})\right)\lambda_{c_i} \quad (1)$$

where n_c is a number of instances in D given c , and instead $P(t_{x_i}|c)$ with $P(t_{x_i}|c,\lambda_{c_i})$ as follows:

$$c(t) = \mathrm{arg}max_{c\in C}\left(p(c)\prod_{i=1}^{m}p(a_i|c,\lambda_{c_i})\right) \quad (2)$$

At last, in that paper, they minimize the cost function to take out a series w_{c_i} for each A_i in class c. The cost function is defined as follows:

$$J(w_c) = \sum_{i=1}^{m}\sum_{a_i}^{A_i}\left(p(a_i|c) - p(a_i|c,w_{c_i})\right)^2 \quad (3)$$

Ⅱ. Our Proposed Approach

As discussion previously, in this section, we propose our novel approach named Naive Bayes Based on Attribute Weighting in Kernel Density Estimation (NBAWKDE). NBAWKDE made a combination which employing conditional mutual information to attribute weighting and kernel density estimation for categorical attribute to relax NB assumption.

2.1 NBAWKDE Kernel Density Estimator

In Equation (1), 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 $\frac{1}{|A_i|}$. We let the bandwidth $\lambda_{c_i} = 1 - w_{c_i}$, the variance with Equation (2) as follows:

$$\kappa(t_{x_i},a_i,w_{c_i}) = \begin{cases} \frac{1}{|A_i|} + \frac{|1-A_i|}{|A_i|}w_{c_i}, \mathrm{if}\ t_{x_i} = a_i \\ \frac{1}{|A_i|}(1-w_{c_i}), \quad otherwise \end{cases} \quad (4)$$

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

$$P(t_{x_i}|c,w_{c_i}) = \frac{1}{n_c}\sum_{i=1}^{n_c}\kappa(t_{x_i},a_i,w_{c_i}) = \frac{1}{|A_i|} + w_{c_i}\left(\overline{f}_c(t_{x_i}) - \frac{1}{|A_i|}\right) \quad (5)$$

NBAWKDE is defined as follows:

$$c(t) = \mathrm{arg}max_{c\in C}\left(p(c)\prod_{i=1}^{m}p(a_i|c,w_{c_i})\right) \quad (6)$$

2.2 NBAWKDE Attribute Weighting

The next question is how can we obtain w_{c_i}? Chang-Hwan et al. [5] generated feature 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_{c_i} for each class c employing conditional mutual information among A_i, A_i and given c. It make sense that if one attribute has less mutual information value with other attributes, this attribute cannot be alternative by other attributes in classification ability, so this attribute will provide more classify ability than other attributes in current class. The average weight $\overline{w_{c_i}}$ of each attribute A_i given c is defined as follows:

$$\overline{w_{c_i}} = 1 - \frac{\sum_{j=1,i\neq j}^{m} I(A_i,A_j|c)}{\sum_{i=1,i\neq j}^{m}\sum_{jj=1,i\neq j}^{m} I(A_i,A_j|c)} \quad (7)$$

where the definition of $I(A_i,A_j|c)$ is follows:

$$I(A_i,A_j|c) = \sum_{i,j}p(a_i,a_j,c)\log\frac{p(a_i,a_j,c)}{p(a_i|c)p(a_j|c)} \quad (8)$$

After linear normalization, w_{c_i} is defined as follows:

$$w_{c_i} = \frac{\overline{w_{c_i}} - \mathrm{min}_{i\in m}(\overline{w_{c_i}})}{\mathrm{max}_{i\in m}(\overline{w_{c_i}}) - \mathrm{min}_{i\in m}(\overline{w_{c_i}})} \quad (9)$$

The output for NBAWKDE is:

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

Ⅲ. Conclusion

In this paper, we proposed a novel approach: Naive Bayes Based on Attribute Weighting in Kernel Density Estimation. NBAWKDE made a combination which employing conditional 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, 21 (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 KDE. (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 (2013)