IEIE Transactions on Smart Processing and Computing, vol. 6, no. 1, February 2017
https://doi.org/10.5573/IEIESPC.2017.6.1.018

IEIE Transactions on Smart Processing and Computing

A Novel Method for a Reliable Classifier using Gradients

Euihwan Han[1] and Hyungtai Cha[2]

[1] Department of Electrical Engineering, Graduate School of Soongsil University / Seoul, Korea
[2] School of Electronic Engineering, Soongsil University / Seoul, Korea

***** Corresponding Author: Hyungtai Cha

Received January 9, 2017; Accepted February 9, 2017; Published February 28, 2017

** Short Paper*

***Abstract*: In this paper, we propose a new classification method to complement a naïve Bayesian classifier. This classifier assumes data distribution to be Gaussian, finds the discriminant function, and derives the decision curve. However, this method does not investigate finding the decision curve in much detail, and there are some minor problems that arise in finding an accurate discriminant function. Our findings also show that this method could produce errors when finding the decision curve. The aim of this study has therefore been to investigate existing problems and suggest a more reliable classification method. To do this, we utilize the gradient to find the decision curve. We then compare/analyze our algorithm with the naïve Bayesian method. Performance evaluation indicates that the average accuracy of our classification method is about 10% higher than naïve Bayes.

***Keywords*: Classifier, Bayes decision, Optimization theory, Gradient

1. Introduction

Recently, there has been increasing interest in pattern recognition in fields such as cognitive science, artificial intelligence, and ergonomics [1]. There are many machine learning algorithms, including Bayesian, neural networks, and support vector machines. Among them, the naïve Bayes classifier assumes the data distribution to be Gaussian and finds the discriminant function using prior probability and likelihood. Finally, it derives the decision curve using the discriminant function [2]. However, this method presumes that all of the features are independent, and thus, errors can occur in the process of obtaining the discriminant function [3]. Therefore, we investigate these problems using a gradient of data distribution, and offer proper solutions that could design a reliable classifier.

2. Related Work

2.1 Naïve Gaussian Bayes Classifier

The naïve Gaussian Bayes classifier assumes the data distribution to be Gaussian, and researchers derive the decision curve using Gaussian distribution. We provide Eqs. (1) and (2) below, and describe this classifier in Fig. 1.

$$N(\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{k/2}\,|\boldsymbol{\Sigma}|^{1/2}} e^{\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)} \tag{1}$$

$$g_i = -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_i) - \frac{1}{2}\log\left(|\boldsymbol{\Sigma}_i|\right) + \log\left(|P(\omega_i)|\right) \tag{2}$$

where $\boldsymbol{\mu}$ is the mean vector, $\boldsymbol{\Sigma}$ is the covariance matrix, and $(\cdot)^T$, $(\cdot)^{-1}$, and $|\cdot|$ are the transpose, inverse, and determinant operator.

3. Troubleshooting

In this section, we discuss the problems with the naïve Bayes classifier. As described in Section 2.1, this classifier assumes that the data distribution is Gaussian and finds the decision curve using a discriminant function (cross-section) of the Gaussian function. However, it is reasonable only when every feature has the same covariance, and features that consist of dimensions are independent [3]. However, this is not often the case in most of the actual data. Besides, errors can occur in cases when researchers simply apply the same confidence level to find the discriminant function. We describe this in Fig. 2.

The larger the magnitude of the gradient, the higher the

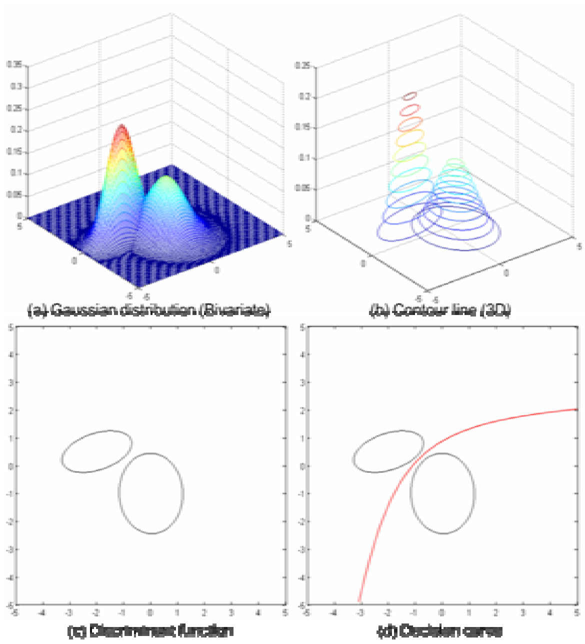


Fig. 1. Naïve Gaussian classifier.

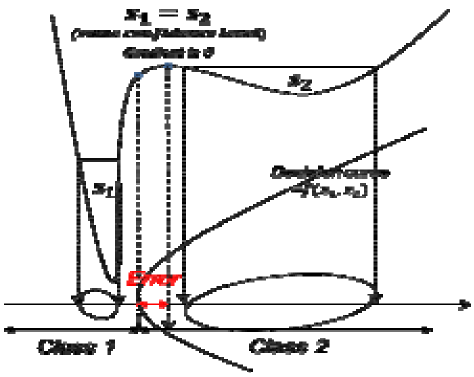


Fig. 2. Troubleshooting.

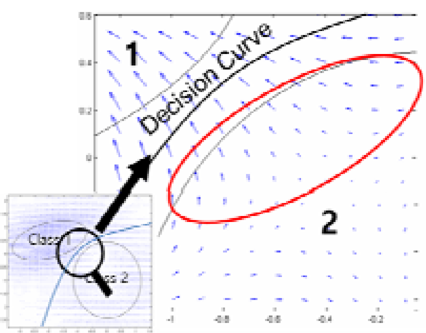


Fig. 3. Gradient of bivariate normal distribution.

data density. The higher the data density, the narrower the cross-section. Therefore, researchers could obtain an accurate decision curve using a gradient line (a set of zero gradient points). We can verify the existing problem by describing it using an illustration. To do this, we utilize two different Gaussian distributions (two-dimension).

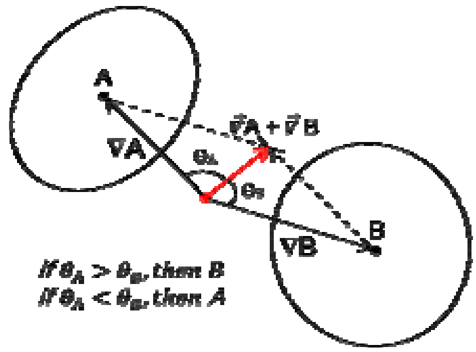


Fig. 4. Decision method using gradient summation.

Using Eqs. (1) and (2), we represent the decision curve and gradient of the data distribution in Fig. 3.

As shown in Fig. 3, we can verify that there are errors in the existing method. We derived the discriminant function and decision curve using Eq. (2), and represent the gradient of each data distribution in Fig. 3. The decision curve is generated between two discriminant functions (the black line in Fig. 4). Based on this decision curve, the lower region is classified as class 2, and the upper area is classified as class 1. However, when we identify data density using the gradient, data that are in class 2 (which was categorized by the naïve Bayesian method) are actually in class 1 (the red circle in Fig. 3). Therefore, in this paper, we propose a new classifier design method to solve the existing problems with using the gradient of data distribution.

4. Algorithm

In this paper, we suggest a novel, reliable classifier method. To do this, we utilized a four-step process: 1. calculate the mean and variance of the data; 2. derive a distribution Equation; 3. find the gradient of the data distribution; and 4. classify class using summation of the gradient vector. For step 1, we utilized the Database for Emotion Analysis using Physiological Signals (DEAP) dataset [4], then proceeded to steps 2 and 3 using Eqs. (1) and (2). Finally, we explain step 4 using Fig. 4. We classify the results using summation of the vector. The direction of the gradient is toward the data distribution's mean value (the Gaussian distribution has extreme value in mean value). And the magnitude of the gradient is proportional to the data density. Therefore, we applied these properties to solve the existing problems. To do this, we calculated the summation of each gradient.

As shown in Fig. 4, we classify the data using the angle (θ_A, θ_B) between the resultant vector ($\vec{\nabla}A + \vec{\nabla}B$) and each gradient vector ($\vec{\nabla}A, \vec{\nabla}B$). If θ_B is smaller than θ_A, it means that the magnitude of $\vec{\nabla}B$ is bigger than that of $\vec{\nabla}A$. Therefore, the point (the red point in Fig. 4) is more affected by distribution *B* than *A*. Consequently, these data are suitable for class *B*. Otherwise, they belong to class *A*. Below, we represent a new decision curve that was applied in our method.

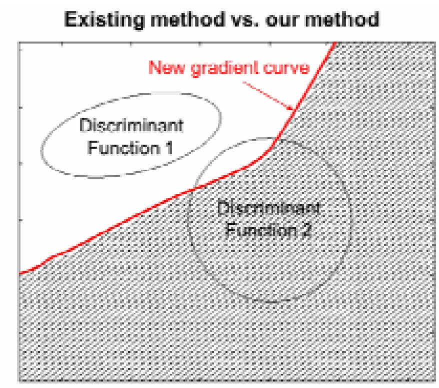


Fig. 5. Gradient curve & contour line.

Table 1. Dataset information.

	Description
Classes	**28 (Videos)**
Features	3 (Arousal, Valence, Dominance)
Scale	9
No. of Respondents	32 (Male: 17/Female: 15) (Mean: 27.1875 / STD: 4.37)
Cross Validation	2 Groups (Randomly Selected)

Table 2. Performance evaluation (naïve Bayes).

2-fold CV	Naïve Bayes Accuracy (%)	Our Algorithm Accuracy (%)
Group 1 (#: 16)	**75**	**85.71**
Group 2 (#: 16)	64.2	71.43
Average	69.6	78.6

5. Performance Evaluation

In this section, we evaluate our method and compare it with a naïve Bayesian classifier. To do this, we utilize the DEAP participant_ratings set [4]. It contains a participant's emotional level (arousal, valence, dominance) about particular videos. We present detailed information about this dataset in Table 1.

The dataset utilizes 28 videos, and respondents check their emotional level pertaining to arousal (the intensity of emotion provoked by a stimulus), valence (the pleasantness of a stimulus), and dominance (the degree of control exerted by a stimulus). Therefore, our training set has 28 classes and three features. We randomly divided the dataset into two groups for cross validation. The performance evaluation of our algorithm and the existing method (naïve Bayes) is shown in Table 2.

6. Conclusion

In this paper, we proposed a new classification method to complement naïve Bayes classifier. To do this we did not utilize the existing discriminant equations, but applied a gradient method. According to the performance evaluation, our classifier has higher accuracy by about 10%, in comparison with naïve Bayes. Consequently, we found that the new method could provide more reliable results. Of course, the gradient method could incur high computation costs, but this lies beyond the scope of the study. Our research focused on designing a classification method with high accuracy. Furthermore, this problem could be offset by downsizing the resolution. In the future, we plan to find ways to reduce computation costs, and will study more reliable classification methods by comparing other pattern recognition algorithms (SVM, NN, etc.).

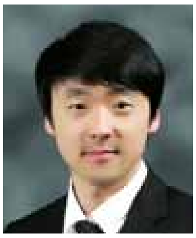
References

[1] J. E. Lee, and S. K. Yoo., "The Design of Feature Selection Classifier based on Physiological Signal for Emotion Detection," *Journal of The Institute of Electronics Engineerings of Korea*, Vol. 50, No. 11, pp. 206–215, 2013. Article (CrossRef Link)

[2] I. S. Oh., *Pattern Recognition*, Seoul: Kyobobook Press. Article (CrossRef Link)

[3] J. D. Rennie, et al., "Tackling the Poor Assumptions of Naive Bayes Text Classifiers," *in Proc. of ICML*, pp. 616–623, 2003. Article (CrossRef Link)

[4] S. Koelstra, et al., "DEAP: A Database for Emotion Analysis using Physiological Signals," *IEEE Transaction on Affective Computing*, Vol. 3, No. 4, pp. 18–31, 2012. Article (CrossRef Link)



Euihwan Han is in Ph.D course in Electronic engineering at the Soongsil University (SSU), Seoul, Korea. He received his B.S. and M.A. degrees in Soongsil University (SSU). His research interests include Pattern recognition, Emotion estimation, Bio-signal and Statistical signal processing. He is a member of the IEIE, KOSES and IEEE.

Hyungtai Cha received his B.S. degree in Electronic Engineering from Soongsil University (SSU), Korea, in 1985. And the M.S and Ph.D. degree from the Department of Electrical Engineering at Pittsburgh University, USA in 1988 and 1993. He was Senior researcher in Samsung Electronics Research Institute, South Korea from 1993 to 1996. He has been chief editor in KOSBE and executive director in KISPS from 2001. Now, he is a full professor at Soongsil University.