Ke GUO, Xia-bi LIU, Lun-hao GUP, Zong-jie LI, Zeng-min GENG, 2018. A new constrained maximum margin approach to discriminative learning of Bayesian classifiers. *Frontiers of Information Technology & Electronic Engineering*, 19(5):639-650. https://doi.org/10.1631/FITEE.1700007

A new constrained maximum margin approach to discriminative learning of Bayesian classifiers

Key words: Discriminative learning; Statistical modeling; Bayesian pattern classifiers; Gaussian mixture models; UCI datasets

Corresponding author: Ke GUO
E-mail: guoke@bit.edu.cn
ORCID: http://orcid.org/0000-0002-9278-4046

Motivations

1. In Bayesian classification, generative learning and discriminative learning are widely used to learn representative class information from samples of different classes.

2. The training data is often insufficient or contains noise, therefore the class distribution estimated by generative learning algorithms often deviates from real ones, leading to unsatisfactory classifiers.

3. Discriminative learning algorithms directly consider the discrimination between classes in the training phase and focus on the difference between classes instead of the distribution of the data within each class.

Main ideas

1. The optimal binary classifier is defined as the one that maximizes the separation between two classes under the constraints of correct classification of training data.

2. The maximum separation between two classes for each sample is pursued by the previous discriminative learning approaches for Bayesian classification.

3. Based on the defined learning criterion, we employ the sequential unconstrained minimization technique .

Methods

1. Learning objective is formatted based on Bayes formula and two constraints are involved .
2. For the optimization, we adopt the exterior point minimization algorithm (Algorithm 1) and gradient descent algorithm (Algorithm 2) to update the model parameters.
3. Two strategies are employed to expand the method to multi-class classification, one-against-one (1v1) and one-against-others (1vO).

Major results

1. Compared with the best previous method, the proposed CMM approach has delivered better recognition rates on 6 test sets out of total 10 sets.

Table 8 Comparisons between recognition rates on test sets from the CMM approach and those of the best results from previous work

Dataset	Previous method	Reference	Recognition rate (%)		IR
			Ours	Best previous	
1	HNB	Jiang et al. (2009)	98.46	95.78	2.80
2	Neural-network	Gorman and Sejnowski (1988)	84.62	90.40	−6.39
3	*k*-SVM	Bredensteiner and Bennett (1999)	52.60	72.43	−27.38
4	SVM	Kwok (1999)	90.48	90.20	0.31
5	3NN	Jiang and Zhou (2004)	97.33	95.67	1.74
6	C5.0	Dvořák and Savický (2007)	78.73	86.77	−9.27
7	WAODE	Jiang et al. (2012)	90.13	95.03	−5.16
8	HNB	Jiang et al. (2009)	86.52	85.46	1.24
9	AODE	Webb et al. (2005)	86.44	84.87	1.85
10	WAODE	Jiang et al. (2012)	98.33	96.63	1.76

IR denotes the increase in rates associated with the constrained maximum margin (CMM) approach, compared with the best results from previous work. HNB: hidden naïve Bayes; SVM: support vector machine; NN: nearest neighbor; AODE: aggregating one-dependence estimators; WAODE: weighted AODE

Major results (Cont'd)

2. Recognition rates on test sets of different datasets, the proposed CMM method has achieved the best average result.

Table 9 Recognition rates on test sets: comparisons between CMM and TAN, AODE, HNB, WAODE, EM, and SVM

Dataset	Recognition rate (%)						
	CMM	TAN	AODE	HNB	WAODE	EM	SVM
1	98.46	95.43	95.25	95.78	95.08	95.08	98.20
2	84.62	75.39	79.04	80.89	78.04	77.88	80.77
3	52.60	58.64	61.13	59.33	59.58	36.82	24.52
4	90.48	81.03	80.54	80.79	80.52	88.24	87.43
5	97.33	94.07	94.47	93.93	95.87	96.00	96.67
6	78.73	82.92	81.13	81.49	81.07	80.80	70.21
7	90.13	93.91	92.94	94.72	95.03	91.21	92.47
8	86.52	78.38	85.22	85.46	85.00	86.36	87.04
9	86.44	79.10	84.87	84.31	84.00	86.08	86.48
10	98.33	93.26	96.07	94.94	96.63	97.22	95.55
Average	86.36	83.21	85.07	85.16	85.08	83.57	81.93

CMM: constrained maximum margin; TAN: tree augmented naïve Bayes; AODE: aggregating one-dependence estimators; HNB: hidden naïve Bayes; WAODE: weighted AODE; EM: expectation maximization; SVM: support vector machine

Major results (Cont'd)

3. The non-parametric Wilcoxon test is performed for comparison of different methods. The CMM approach significantly outperforms the EM and SVM method.

Table 10 Ranks computed by the Wilcoxon test

Algorithm	Sum of ranks for the datasets						
	CMM	TAN	AODE	HNB	WAODE	EM	SVM
CMM	–	42.0	37.0	34.0	35.0	46.0	46.0
TAN	13.0	–	13.0	9.0	9.0	19.0	20.0
AODE	18.0	42.0	–	27.0	32.0	26.0	22.0
HNB	21.0	46.0	28.0	–	33.5	28.0	25.0
WAODE	20.0	46.0	23.0	21.5	–	21.0	26.0
EM	9.0	36.0	29.0	27.0	24.0	–	29.0
SVM	9.0	35.0	33.0	30.0	29.0	26.0	–

Values above the diagonal indicate that the algorithm in the row outperforms the one in the column; values below the diagonal indicate the opposite. CMM: constrained maximum margin; TAN: tree augmented naïve Bayes; AODE: aggregating one-dependence estimators; HNB: hidden naïve Bayes; WAODE: weighted AODE; EM: expectation maximization; SVM: support vector machine

Conclusions

1. The separation between all the samples of a class and those of another class would be maximized, since the margin was defined on the class level.

2. The penalty factor and its magnification coefficient of the constraint item would boost the learning results by increasing the effect of misclassification samples during the optimization.

3. Experimental results demonstrated that our CMM approach was effective and promising.