Modern Education
and Computer Science
PRESS

A New Method of Signature Verification Based on Biomimetic Pattern Recognition Theory

Yan Wu, Hui Geng, Xiao-yue Bian

Department of Computer Science and Technology, Tongji University, Shanghai, China

Abstract

Aim at the difficulty and low recognition rate of signature verification, this paper introduces biomimetic pattern recognition theory and applies it to the problem. According to the features of the signature samples, the coverage in the high-dimension feature space is built, one class of samples are all covered with a super-sausage neuron chain. As the radius selection of the super-sausage neurons maybe unreasonable, unwanted area may be covered and correct recognition rate will reduce. So this paper uses the relationship of the distance between the two training samples and the average distance of all the neurons to adjust the radius of the super-sausage neuron automatically. Finally, the experiments show that compared to traditional pattern recognition method, biomimetic pattern recognition theory used in signature verification have a better recognition result and is more effective.

Index Terms: Signature verification; biomimetics pattern recognition; high-dimension space; super-sausage neuron

1. Introduction

In our life, hand-written signature verification is the most familiar person identification method which based on hand-written features, it's also highly accepted. In the 1960s, A.J.Mancerj made a report "the analysis of the feasibility of hand-written signature verification used in person identification", which told us the possibility of hand-written signature verification used in person identification. Now, the research of signature verification most focus on off-line recognition, and there are a few algorithms, such as verifying with the combination of discrete HMM[1]; simulating the signature path with B ézier curve, extracting the hand-written features and searching the optimal solution with multi-sufficiency function genetic algorithm[2]; extracting the profile features with the hand-written direction of the intersection of the strokes and comparing the random forged signatures using neural network or SVM[3]; extracting the inflection features of the curve with wavelet transformation[4]; recognition methods based on moments and evidence theory[5]; recognition methods based on wavelet transformation[6] and

* Corresponding author.
E-mail address: yanwu@tongji.edu.cn

so on. From above, we can see that they are all improvement or optimization based on existing algorithms. But there're still difficulties when doing recognition and the correct recognition rate is low. This paper will apply biomimetic pattern recognition(BPR) theory to this problem and discuss the effect to the coverage when the neuron's radius change. Experiments are done to compare traditional pattern recognition method with BPR and identify the accuracy and feasibility of BPR.

2.   Biomimetic Pattern Recognition

BPR[7] was a new pattern recognition theory and method proposed by academician Wang Shoujue in 2002. Traditional pattern recognition methods focus their attention on "discrimination", while BPR focus on "cognition". BPR's basic starting point is to cognize the samples one class by one, not to discriminate the samples to many classes. According to topology, BPR builds coverage in the high-dimension space with the geometry theoretical knowledge and image thinking and then achieves it with multi-weight neural network or neural computer[7][8].

BPR's basic point is that it confirms the continuity of similar samples in the feature space[7]. It means "In nature, If there're two 'homologous' things which are similar with the thing desired to be known, and the differences between these two things are gradual or non-quantized, there must be at least a gradual process between the two similar things. The things in the gradual process are all belong to the same class", described as a collection of mathematical formulas:

In the feature space F^n, we suppose that set A is a point set including all samples in class A. There are two random items x and y in it , when ε is any value greater than 0, there must exist an set B :

$$B = \{x_1, x_2, ..., x_n \mid x_1 = x, x_n = y, n \in N,$$
$$\rho(x_m, x_{m+1}) < \varepsilon, \varepsilon > 0, 1 \le m \le n-1, m \in N\}, B \subset A \tag{1}$$

$\rho(x,y)$ means the distance between the two samples. In the feature space, the set of points which contains the continuous mapping of the samples belong to one class is made as a closed set. This closed set is the coverage of this class in the feature space, all samples belong to this class is in the coverage.

In the real BPR, we need to judge an item belong to the set Pa or not,

$$P_a = \{x \mid \rho(x, y) \le k, y \in A, x \in F^n\} \tag{2}$$

where k is a distance threshold. We should build a n-dimension coverage to cover the set Pa in the feature space. The coverage is like that: the union of the infinitely many n-dimensional hyperspace with infinite number of points in the 'manifold' of different dimensions as the globes and the constant k as the radius. We can also describe it as the product topology of the set A and the n-dimension hyperspace. The task of BPR is to judge whether the mapping of the "recognized sample" in the feature space is in the set Pa or not.

3.   Bpr Applied to Signature Verification

A. Preprocessing and Feature Extraction

During the signature verification, first we have to preprocess the original images. Preprocessing is the first step in signature verification, it's also a crucial step. The result will directly affect the recognition rate of the verification. In off-line signature verification, the samples must be recognized by optical scanning equipment first, so many kinds of noise appear. In this paper, smooth noise reduction, binarization and normalization are done to the images to reduce the effect of noise on the feature extraction and signature verification.

In the feature extraction stage, after preprocessing the images, we extract the texture features, shape features, pseudo-dynamic features and pseudo-Zernike moments features. In texture features, we use autocorrelation function to extract the roughness[9], tilt vector feature and stroke density feature of the image. Shape features include the number of connected components, aspect ratio and effective aspect ratio. With Laplacian, we extracted the high gray regional features of the signature images. Including the previous 10-dimension features of the pseudo-Zernike moments, we extract 24-dimension features totally.

B. The Signature Verification Model Based on BPR

This paper is about off-line signature verification, in this problem, signatures written in different time are not the same, but they are similar. And the similarity is the basis of signature verification. Fig. 1 is the signature sequence written in different time, and you can see the similarities of all signatures.

Fig 1. Signature sequence written in different time
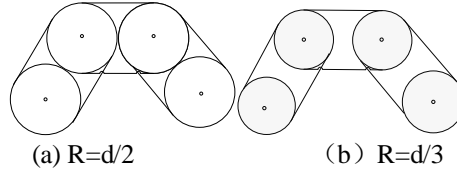

(a) R=d/2   （b）R=d/3

Fig 2. The effect to the coverage when the radius changes

Signature is the writing habit formed in a person's daily life, they are similar and also mutative. Signatures in different time are similar, and the similarity can be seen as the continuity between different signatures. With the change of time, the difference between signatures can be seen as the volatility around the original writing habit. So with the change of one dimension (time), the change of the mapping-which is mapped into the high-dimension feature space is also continuous. According to the BPR theory-the continuity of similar samples in the feature space[7], the set of the corresponding mapping should distribute as one-dimension manifold, while the change of other direction is tiny. We can say that the coverage's shape of one class in the feature space is the product topology of a one-dimension manifold homeomorphic with the curve segment and a 24-dimensional hyperspace, and then a closed subspace of this class of samples is produced.

Assuming the curve segment to be A, the radius of the hyperspace to be R, the subspace of this class-Pa should be：

$$P_a = \{x \mid \min(\ \rho(x,y)) < R, y \in A, x \in F^{24}\}$$

(3)

Assuming the number of training samples of each person's signature to be N, and the training set is $S = \{x \mid x = s_1, s_2, ..., s_N\}$. S1,S2,...SN are different signatures of one person in different time.

Since the inherent properties of artificial neural network, it's a very appropriate means to realize BPR. In order to use a number of neurons to cover Pa , we can use a few line segment to approximate curve A. Each neuron covers the product topology of a line segment and a 24-dimensional sphere, the coverage Pb is the subspace of that class of samples we get at last. Because the number of training samples is N, we choose N-1 line segment to approximate the curve A, the lines are described as $B_i (i = 1, 2, ..., N-1)$,

$$B_i = \{x \mid x = as_i + (1-a)s_{i+1}, a \in [0,1], s_i \in S, x \in F^{24}\} \tag{4}$$

The coverage of each neuron is

$$P_i = \{x \mid \min(\rho(x,y)) \le B_i, x \in F^{24}\} \tag{5}$$

The coverage of all the neurons (N-1 at most) is

$$P_b = \bigcup_{i=1}^{N-1} P_i \tag{6}$$

The Selection of the Radius of the Super-sausage-neurons

The selection method of the super-sausage neuron's radius was not explicitly given when academician Wang proposed BPR. This paper decides the radius with the distance of two samples of one super-sausage neuron, and adaptively adjusts the radius with the distance between two training samples.

For the super-sausage neuron decided by two samples, The coverage in feature space is different when the radius changes. Fig. 2 shows the effect to the covered area when the radius changes. The Euclidean distance between two samples is d, radius of neurons is R.

From Fig. 2, we can see that the coverage in Fig. 2(a) is bigger than that in Fig. 2(b). So the area the neuron covers is proportional to the radius of the neuron. But it doesn't mean the bigger the radius is, the better the coverage is. It depends on the practical application. For signature verification, when the radius grows, the wrong recognition rate of a fake signature will increase, so the safety factor will reduce. On the contrary, if the radius is very small, the correct recognition rate of a genuine signature will reduce. Although the safety factor increases, the wrong recognition rate of the genuine signature will increase. In order to obtain a certain precision of signature, the neuron's radius can be changed within a certain range.

This paper proposes a simple adjustment method through some experimental data. When the distance between two samples of one neuron is bigger than the average distance of all training samples, it means the two samples are far from each other and the coverage is big. At that time, we will adjust the radius to a quarter of the distance between the two samples. When the distance between two samples of one neuron is smaller than the average distance of all training samples, it means the two samples are close to each other and the coverage is small. At that time, we should choose bigger radius and form bigger coverage. As shown in (7):

$$R = \begin{cases} \dfrac{d}{4}, d > avg\_d \\ \dfrac{d}{2}, d \le avg\_d \end{cases} \tag{7}$$

where avg_d is the average distance of all the trained samples.

C. Covering Algorithm

BPR needn't training when doing recognition, the process of recognizing is also the process of training. Any two samples in the set of training samples can produce a super-sausage neuron, and the neuron forms a closed set in the feature space, then it becomes a coverage. The coverage of all the super-sausage neurons forms the coverage of a class of things. We assume the sample set of one class of things to be S, and the number of samples is N, the covering process is as below:

Step 1:Find two samples that their distance is the smallest in the S, and use them to form the first super-sausage neuron. Then we put the two samples into the S'.

Step 2:Select a random sample A from the rest samples in the S, and find out whether it's contained in the coverage of the existed super-sausage neurons. If be, remove A from S, and put it into S'. If not, turn to step 3.

Step 3:Find sample B who is most near sample A from S', and use A and B to form a new super-sausage neuron. Then calculate the radius of this neuron with formula 7 and add the sample B to S'.

Step 4:Find out whether there are still samples in S or not. If be, turn to step 2; if not, turn to the end.

When we do the recognition, we only need to find out whether the sample is in the coverage of all the super-sausage neurons. If it is, we consider it genuine signature, otherwise we refuse to recognize.

4. Experiment Results and Analysis

This experiment collected 900 signatures of 15 individuals, 60 each person, including 40 genuine signatures and 20 fake ones. To verify the effectiveness of the method, our experiments selected some genuine signatures as training samples and the rest(including genuine and fake signatures) as the test samples. Tab. 1 shows the average result of our 10 experiments for different radius. In these experiments, we selected 12 random genuine signatures of each person as the training samples.

Table 1 experiment result of BPR for different radius

		R =d/4	R =d/3	R =d/2	This paper's method
Genuine signatures	Correct recognition rate	89.1%	90.5%	92.6%	94.5%
	False acceptance rate	3.6%	4.3%	4.8%	3.2%
	Rejection rate	7.3%	5.2%	2.6%	2.3%
Fake signatures	Correct recognition rate	90.9%	92.1%	93.2%	95.2%
	False acceptance rate	3.8%	4.3%	4.6%	2.2%
	Rejection rate	5.3%	3.6%	2.2%	2.6%

From Tab. 1, we know when the radius R is d/4, the correct recognition rate is lowest, and because the radius is small, the false acceptance rate is also lowest, and at the same time, the rejection rate is highest. When the radius R is d/2, the correct recognition rate increases a lot. At the same time, because the radius grows, the false acceptance rate increases and the rejection rate reduces. However, in the method our paper proposed, the selection of radius is relative with the training samples, it gives attention to the coverage of every super-sausage neuron. So the correct recognition rate is high, at the same time, the false acceptance rate and the rejection rate is also acceptable. Thus we can demonstrate the effectiveness of the proposed method.

The signatures of one person are relatively stable, when judging a test sample is the person's genuine signature or not, we only need to find the relationship between the feature of the test sample and the feature of that person's genuine signatures (the set of training samples). We can say the signature verification problem is actually a two-class classification problem. So we compare our method(BPR) with KNN and SVM. KNN gets the recognition result through the relationship between the input sample and the k nearest samples of that class of training samples. In SVM, everyone's signature verification problem is a two-class classification problem, and it judges the test samples' true or false by the division of the input sample and this class of training samples. In

KNN, we give k the value 13 and 8. In SVM, we use Polynomial kernel function(SVM(1)) and Gaussian kernel function(SVM(2)), in the Polynomial kernel function, d=2,c=1; in the Gaussian kernel function, σ=2. Tab. 2 shows the average result of 10 experiments for different quantity of test samples. Because KNN and SVM only give out an optimal solution, not the correct solution, so the rejection rates of those two algorithms are 0.

Table 2 comparison of recognition rate

Number of samples	Recognition methods	Genuine signature			Fake signature			Time(s)
		Correct recognition rate	False acceptance rate	Rejection rate	Correct recognition rate	False acceptance rate	Rejection rate	
25	KNN	89.4%	10.6%	0	91.6%	8.4%	0	2.24
	SVM (1)	92.2%	7.8%	0	95.3%	4.7%	0	3.16
	SVM (2)	94.9%	5.1%	0	95.7%	4.3%	0	3.25
	BPR	95.2%	1.9%	2.9%	96.3%	1.6%	2.1%	2.28
15	KNN	84.3%	15.7%	0	86.8%	13.2%	0	1.73
	SVM (1)	87.4%	12.6%	0	88.7%	11.3%	0	1.97
	SVM (2)	89.5%	10.5%	0	90.4%	9.6%	0	2.13
	BPR	94.9%	2.2%	2.9%	95.7%	2.1%	2.3%	1.82

From Tab. 2, we can see the algorithm we proposed has absolute advantage in time and efficiency. Time we consumed is close to KNN, but our correct recognition rate is higher and false acceptance rate is lower. KNN must sort the training samples, so the more the samples are, the more time is consumed. Under the same experiment condition, our correct recognition rate has litter difference with SVM, but SVM need more time, about 1 second more in our experiment. BPR doesn't need to train, it builds the coverage in the high-dimension space, so it can recognize directly and save time. From the analysis above, we know BPR improve a lot from traditional pattern recognition method.

Additional speaking, when the quantity of the training samples is small, BPR works better than other algorithms. As the quantity declines, the correct recognition rate of the traditional recognition method will decrease obviously. While the recognition rate of BPR will not change a lot. Needn't many training samples, BPR can build the space coverage of one class of training samples, so it is suitable for the recognition of small sample set.

5.   Conclusion

By analyzing the basic principles and theory of biomimetic pattern recognition, we use it to build the coverage of super-sausage neurons and apply it to the signature verification problem. This paper proposed a method to adjust the radius of the super-sausage neurons adaptively. At last, we use experiments to prove that the correct recognition rate of BPR is better than traditional pattern recognition methods.

References

[1]Luana Batista, Eric Granger, Robert Sabourin. A multi-hypothesis approach for off-line signature verification with HMMs[J]. 2009 10th International Conference on Document Analysis and Recognition, IEEE:13115-1319

[2]Diego Bertolini, Luiz S. Oliveira, Edson Justino, et al. Ensemble of classifiers for off-line signature verification [J]. 2008 IEEE International Conference on Systems, Man and Cybernetics (SMC 2008),IEEE:283-288

[3]Vu Nguyen, Michael Blumenstein, Vallipuram Muthukkumarasam, et al. Off-line signature verification using enhanced modified direction features in conjunction with neural classifiers and support vector machines. Ninth International Conference on Document Analysis and Recognition, IEEE, 2007, Vol.2(9):734-738

[4]Xiufen Ye, Weiping Hou, Weixing Feng. Off-line handwritten signature verification with inflections feature. International Conference on Mechatronics and Automation, Niagara Falls, Canada, IEEE, 2005.7:787-792

[5]Gang Chen, Bicheng Li, Jun Zhuang. Off-line handwritten signature verification based on moments and evidence theory[J] (in Chinese). Computer Applications, 2005, 18(18):57-60

[6]Cong Shen. Signature verification based on advanced multi_channel Gabor-wavelet transformation[D] (in Chinese). Master thesis, Beijing University of Technology, 2002

[7]Shoujue Wang. Biomimetic pattern recognition [J] (in Chinese), Acta Electronica Sinica,2002, 30(10):1417-1420

[8]Shoujue Wang. First step to multi-dimensional space biomimetic infomatics[M] (in Chinese). Beijing: National Defence Industry Press, 2008.1

[9]Bicheng Li, Tianqiang Peng, Bo Peng. Intelligent image processing[M] (in Chinese). Beijing:Electronic Industry Press, 2004