Neurocomputing 151 (2015) 1468–1476

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Multi-hypothesis nearest-neighbor classifier based on class-conditional weighted distance metric

Lianmeng Jiao [a,b], Quan Pan [a,*], Xiaoxue Feng [a,c]

[a] School of Automation, Northwestern Polytechnical University, 710072 Xi'an, PR China
[b] UMR CNRS 7253, Heudiasyc, Université de Technologie de Compiègne, 60205 Compiègne, France
[c] UMR CNRS 6279, ICD-LM2S, Université de Technologie de Troyes, 10010 Troyes, France

ARTICLE INFO

Article history:
Received 8 February 2014
Received in revised form
24 May 2014
Accepted 13 October 2014
Communicated by Xianbin Cao
Available online 28 October 2014

Keywords:
Pattern classification
Weighted distance metric
Multi-hypothesis nearest-neighbor classifier
Dempster–Shafer theory

ABSTRACT

The performance of nearest-neighbor (NN) classifiers is known to be very sensitive to the distance metric used in classifying a query pattern, especially in scarce-prototype cases. In this paper, a class-conditional weighted (CCW) distance metric related to both the class labels of the prototypes and the query patterns is proposed. Compared with the existing distance metrics, the proposed metric provides more flexibility to design the feature weights so that the local specifics in feature space can be well characterized. Based on the proposed CCW distance metric, a multi-hypothesis nearest-neighbor (MHNN) classifier is developed. The scheme of the proposed MHNN classifier is to classify the query pattern under multiple hypotheses in which the nearest-neighbor sub-classifiers can be implemented based on the CCW distance metric. Then the classification results of multiple sub-classifiers are combined to get the final result. Under this general scheme, a specific realization of the MHNN classifier is developed within the framework of Dempster–Shafer theory due to its good capability of representing and combining uncertain information. Two experiments based on synthetic and real data sets were carried out to show the effectiveness of the proposed technique.

1. Introduction

The nearest-neighbor (NN) rule, first proposed by Fix and Hodges [1], is one of the most popular and successful pattern classification techniques. Given a set of N labeled samples (or prototypes) $T=\{(\mathbf{x}^{(1)},\omega^{(1)}),\dots,(\mathbf{x}^{(N)},\omega^{(N)})\}$ with input vector $\mathbf{x}^{(i)}\in\mathbb{R}^D$ and class label $\omega^{(i)}\in\{\omega_1,\dots,\omega_M\}$, the NN rule classifies a query pattern $\mathbf{y}\in\mathbb{R}^D$ to the class of its nearest neighbor in the training set T. The basic rationale of the NN rule is both simple and intuitive: patterns close in feature space are likely to belong to the same class. The good behavior of the NN rule with unbounded numbers of prototypes is well known [2]. However, in many practical pattern classification applications, only a small number of prototypes are available. Typically, under such a scarce-prototype framework, the ideal asymptotical behavior of the NN classifier degrades dramatically [3]. This problem has driven the growing interest in finding variants of the NN rule and adequate distance measures (or metrics) that help improve the NN classification performance in small data set situations.

As the core of the NN rule, the distance metric plays a crucial role in determining the classification performance. To overcome the limitations of the original Euclidean (L2) distance metric, a number of adaptive methods have recently been proposed to address the distance metric learning issue. According to the structure of the metric, these methods can be mainly divided into two categories: global distance metric learning and local distance metric learning [4]. The first learns the distance metric in a global sense, i.e., to share the same simple weighted (SW) distance metric for all of the prototypes:

$$d_{SW}(\mathbf{x},\mathbf{y})=\sqrt{\sum_{j=1}^{D}\lambda_j^2(x_j-y_j)^2}, \tag{1}$$

where $\mathbf{x}$ is a prototype in the training set, $\mathbf{y}$ is a query pattern to be classified, and λ_j is the weight of the j-th feature. Based on the above distance metric, the feature weights learning in [5,6] is formulated as a linear programming problem that minimizes the distance between the data pairs within the same classes subject to the constraint that the data pairs in different classes are well separated. Eick et al. [7] introduce an approach to learn the feature weights that maximize the clustering accuracy of objects in the

* Corresponding author at: School of Automation, Northwestern Polytechnical University, 710072 Xi'an, PR China Tel.: +86 29 88431307; fax: +86 29 88431306.
E-mail addresses: lianmeng.jiao@utc.fr (L. Jiao), panquannpu@gmail.com (Q. Pan), xiaoxue.feng@utt.fr (X. Feng).

http://dx.doi.org/10.1016/j.neucom.2014.10.039
0925-2312/

training set, and similarly, the classification error rate of objects in the training set is employed to evaluate the feature weights in [8]. Although the above global distance metric learning methods are intuitively correct, they are too coarse, as the feature weights of the distance metric are irrelevant with the prior-known class labels of the prototypes. This issue becomes more severe when some features behave distinctly for different classes (for example, one feature may be more discriminative for some classes but less relevant for others) [9]. Thus, many methods [10–14] have been developed to learn a distance metric in a local setting, i.e., the feature weights may be different for different prototypes. The most representative method is the class-dependent weighted (CDW) distance metric proposed by Paredes and Vidal [15,16], which is related to the class index of the prototype:

$$d_{CDW}(\mathbf{x},\mathbf{y})=\sqrt{\sum_{j=1}^{D}\lambda_{c,j}^2(x_j-y_j)^2}, \tag{2}$$

where c is the class index of prototype $\mathbf{x}$. Although the above CDW distance metric provides more freedom than the SW metric, the following example illustrates that this distance metric is insufficient to reflect the local specifics in feature space for query patterns in different classes. Fig. 1 illustrates a simple three-class classification problem, where the data in each class are uniformly distributed. $(\mathbf{x}^{(1)},A)$, $(\mathbf{x}^{(2)},B)$ and $(\mathbf{x}^{(3)},C)$ are two-dimensional data points in training set T. $\mathbf{y}_1$ and $\mathbf{y}_2$ are the query data to be classified. Considering the classification of data $\mathbf{y}_1$, when calculating the distance between $\mathbf{x}^{(2)}$ and $\mathbf{y}_1$, intuitively, to avoid classifying it to Class B mistakenly, the feature value in the X-axis should be given a larger weight. However, in classifying data $\mathbf{y}_2$, it is reasonable that the feature value in the Y-axis should be given a larger weight to determine the distance between $\mathbf{x}^{(2)}$ and $\mathbf{y}_2$. That is, the feature weights should also be related to the class labels of the query patterns to be classified.

Motivated by the above consideration, in this paper we propose a more general distance metric that associates with both the class labels of the prototypes and the query patterns. As in classification problems the class label of the query pattern is not prior-known, this general distance metric only makes sense when conditioned on the assumption that the query pattern belongs to a specified class. Therefore, we define this type of variant as a class-conditional weighted (CCW) distance metric. Compared with the existing distance metrics mentioned above, the CCW metric provides more flexibility to design the feature weights so that the local specifics in feature space can be well characterized.

Based on the CCW distance metric, this paper develops a multi-hypothesis nearest-neighbor (MHNN) classifier. The main scheme

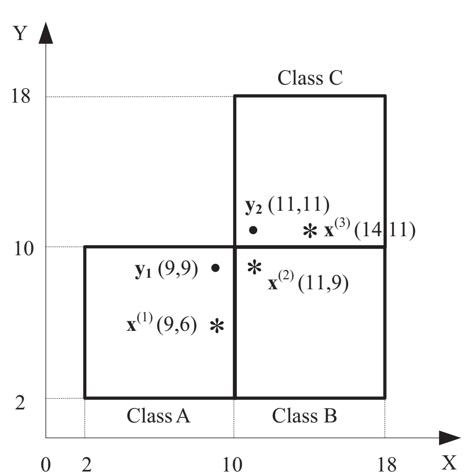

Fig. 1. A three-class classification example.

of this method is to classify the query pattern under multiple hypotheses in which the nearest-neighbor sub-classifiers can be implemented based on the CCW distance metric and then to combine the classification results of multiple sub-classifiers to obtain the final result. A variety of schemes have been proposed for deriving a combined decision from individual decisions, such as majority voting [17], Bayes combination [18], multilayered perceptrons [19], and the Dempster–Shafer theory (DST) [20–22]. In this paper, a specific realization of the MHNN classifier is developed within the framework of DST due to its good capability of representing and combining uncertain information which is always encountered in classification problems.

The rest of this paper is organized as follows. In Section 2, the class-conditional weighted distance metric is defined and then both a heuristic method and a parameter optimization procedure are designed to derive the involved feature weights. The multi-hypothesis nearest-neighbor classifier is designed and realized within the framework of DST in Section 3. Two experiments are given to evaluate the performance of the proposed method in Section 4. Finally, Section 5 concludes the paper.

2. Class-conditional weighted distance metric

2.1. Definition

Before defining the class-conditional weighted distance metric for the purpose of classification, we will first give a general weighted distance metric between two patterns with prior-known class labels as follows.

Definition 1 (*General weighted distance metric*). Suppose $\mathbf{x}^{(m)}$ and $\mathbf{x}^{(n)}$ are two D-dimensional patterns with class labels ω_p and ω_q. A general weighted distance metric between $\mathbf{x}^{(m)}$ and $\mathbf{x}^{(n)}$ can be defined as

$$d(\mathbf{x}^{(m)},\mathbf{x}^{(n)})=\sqrt{\sum_{j=1}^{D}\lambda_{p,q,j}^2(x_j^{(m)}-x_j^{(n)})^2}, \tag{3}$$

where $\lambda_{p,q,j}$ is a constant that weights the role of the j-th feature in the distance metric between class ω_p and class ω_q.

This definition includes, as particular cases, the distance metrics revisited in the Introduction. If $\lambda_{p,q,j}=1$ for all $p=1,\ldots,M$, $q=1,\ldots,M$, $j=1,\ldots,D$, the above distance metric is just the L2 distance metric. Moreover, the SW and CDW distance metrics correspond to the cases where the metric weights are not relevant to the class labels or are only dependent on the class label of the first pattern, respectively. Therefore, the above weighted distance metric provides a more general dissimilarity measure than the L2, SW or CDW metrics because the weights depend on both class labels of the two considered patterns.

In NN-based classification, the problem is to calculate the distance between a prototype and a query pattern, while the class label of the latter is not prior-known. So, for the purpose of classification, the above general distance metric only makes sense conditioned on the assumption that the query pattern belongs to a specified class ω_q. We will define this type of variant as follows.

Definition 2 (*Class-conditional weighted distance metric*). Let $T=\{(\mathbf{x}^{(1)},\omega^{(1)}),\ldots,(\mathbf{x}^{(N)},\omega^{(N)})\}$ be a set of prototypes. The class-conditional weighted (CCW) distance metric between a query pattern $\mathbf{y}$ and a prototype $\mathbf{x}\in T$ can be defined as

$$d_{CCW}(\mathbf{x},\mathbf{y})=\sqrt{\sum_{j=1}^{D}\lambda_{p,q,j}^2(x_j-y_j)^2}, \tag{4}$$

where p is the class index of the prototype $\mathbf{x}$ and q is the hypothesized class index of the query pattern $\mathbf{y}$.