Pattern Recognition 64 (2017) 268–276

Contents lists available at ScienceDirect

Pattern Recognition

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

Efficient temporal pattern recognition by means of dissimilarity space embedding with discriminative prototypes

Brian Kenji Iwana[a,b,*], Volkmar Frinken[a,c,d], Kaspar Riesen[e], Seiichi Uchida[a]

[a] Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan
[b] Institute of Decision Science for a Sustainable Society, Kyushu University, Fukuoka, Japan
[c] Electrical and Computer Engineering, UC Davis, Davis, CA, USA
[d] Onai Technology, San Jose, CA, USA
[e] University of Applied Sciences and Arts Northwestern Switzerland, Olten, Switzerland

ARTICLE INFO

Keywords:
Temporal patterns
Online digit classification
Dissimilarity representation
Ensemble classification
Dissimilarity space embedding

ABSTRACT

Dissimilarity space embedding (DSE) presents a method of representing data as vectors of dissimilarities. This representation is interesting for its ability to use a dissimilarity measure to embed various patterns (e.g. graph patterns with different topology and temporal patterns with different lengths) into a vector space. The method proposed in this paper uses a dynamic time warping (DTW) based DSE for the purpose of the classification of massive sets of temporal patterns. However, using large data sets introduces the problem of requiring a high computational cost. To address this, we consider a prototype selection approach. A vector space created by DSE offers us the ability to treat its independent dimensions as features allowing for the use of feature selection. The proposed method exploits this and reduces the number of prototypes required for accurate classification. To validate the proposed method we use two-class classification on a data set of handwritten on-line numerical digits. We show that by using DSE with ensemble classification, high accuracy classification is possible with very few prototypes.

1. Introduction

An important issue in pattern recognition is the method in which the data is represented. Using a representation that best describes the data or that captures the discriminating features is one of the most important factors in a successful machine learning model [1]. For that reason, much research is dedicated to the preprocessing, feature design, and transformation of data [2,3]. In general, there are two major approaches to this problem, statistical and structural. A statistical approach uses feature vectors as basic representation formalism, while the latter employs symbolic data structures (e.g. strings, trees, or graphs). Whether we use feature vectors or symbolic data structures for pattern representation, the goal is to find a good representation of the data with the foundation to support machine learning.

Dissimilarity representation is a novel approach in pattern recognition [4] in which the data is represented as the difference, or *dissimilarity*, to other objects or patterns. Instead of using a pattern's features, data representation is done based on the patterns relations to other patterns. In this manner, we represent the data as vectors of dissimilarities rather than feature vectors and are able to embed the dissimilarity representation into vector space, or dissimilarity space embedding (DSE). The intuition of this approach is the notion that distance between patterns of similar categories is smaller than patterns of different categories. This fundamental view of patterns as distances also grants the to use complex patterns, like symbolic data structures, with the mathematical and machine learning foundation of vector space [5]. By embedding the patterns into dissimilarity space, we can bridge the gap between complex pattern recognition and feature vector recognition.

This paper specifically tackles temporal pattern recognition. Temporal patterns present an aforementioned complex pattern due to the interconnectivity and dependency on sequence order. Also, when learning temporal patterns, temporal distortions and length variations must be considered. These obstacles are not addressed by classical fixed-dimensional Euclidean pattern recognition. Rather than using a complex model to represent temporal patterns, we can accuracy represent them using DSE and still use the tools granted by vector space. To accomplish this, we embed the temporal patterns into a dissimilarity space by using dynamic time warping (DTW) as a dissimilarity measure. DSE is an N-dimensional feature vector who elements are the DTW distance between prototype patterns and sample patterns. The N-dimensional space is now dimensionally fixed to the

* Corresponding author at: Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan.
E-mail addresses: brian@human.ait.kyushu-u.ac.jp (B.K. Iwana), kaspar.riesen@fhnw.ch (K. Riesen), uchida@ait.kyushu-u.ac.jp (S. Uchida).

http://dx.doi.org/10.1016/j.patcog.2016.11.013
Received 9 January 2016; Received in revised form 17 September 2016; Accepted 16 November 2016
Available online 21 November 2016
0031-3203/

number of prototypes N and can be used for the analysis of temporal patterns.

Recently, it has been a trend in pattern recognition that by increasing the data set size, a higher accuracy could be achieved. Banko and Brill [6] demonstrated that for natural language processing, the number of words was more important than the complexity of the classifier. The work of 80 million tiny images [7] has shown that even with a simple nearest neighbor classification performs sufficiently with a massive number of image patterns. Similar trials with massive patterns has been made for image segmentation [8], video image recognition [9,10], and handwritten character images [11], proving the value of using massive pattern numbers. Therefore, intuitively, to exploit DSE to its potential, we need to use large data.

However, by increasing the size of the data set, you also increase the computational requirement. When using massive pattern numbers, you can encounter the problem of having a huge computational requirement. Typically, there are three primary ways to address this issue: reducing the size of the patterns or features, using efficient search schemes, and selecting meaningful patterns. Tiny images [7] and tiny videos [9] overcome the computational requirement by reducing the image resolution. Other solutions include creating low-dimensional representations via feature selection [12,13]. Another solution to the computational problem is to use optimized search schemes such as estimation, approximate nearest neighbor [14], and vocabulary tree search [15]. A third solution is to select only the meaningful patterns or prototypes useful for the classifier. Classical methods include prototype selection, edition, and condensation [16].

Besides the problem of computational time when using a massive data set and a large set of prototypes, we are also creating a high dimensional vector space that is subject to the *curse of dimensionality* [17]. Therefore, we need to reduce the dimensionality of the DSE. AdaBoost offers two advantages for this purpose: (1) building a strong ensemble classifier from weak learners and (2) being used for selecting discriminative features when each weak learner depends on single features [18]. To solve the computational challenge, we apply AdaBoost to select the meaningful features and in this the case of DSE, prototypes as references to dissimilarities to samples. The end result is a novel, optimized prototype selector scheme for DSE.

The contribution of this paper is to address the massive pattern recognition of temporal patterns. Firstly, we embed temporal patterns into a vector space using a DTW-based dissimilarity measure. This allows us to represent temporal patterns as vectors of dissimilarities to prototypes. Secondly, we develop a new method of prototype selection by using AdaBoost as a feature selector for the resulting DSE. We can increase the computational efficiency by only using the prototypes within the ensemble classifier created by AdaBoost. Finally, we show that the accuracy of the classifier constructed by AdaBoost can be improved by preparing the prototype pool with patterns of classes external to a two-class classification.

The remaining of this paper is organized as follows. Section 2 reviews the related work in prototype selection and DSE. Section 3 elaborates on dissimilarity representation and the relation to vector space embedding. Section 4 provides a more detailed description of the method proposed. In Section 5, we confirm that the proposed method and discuss the usability of patterns of classes external to the two-class classifier. Finally, Section 6 draws a conclusion and describes future work.

2. Related works

In this section, we will briefly discuss prototype selection and provide the previous works related to DSE.

2.1. Prototype selection

The goal of prototype selection is to reduce an initial set of prototypes to a subset while retaining as much information as possible. There are many aspects and types of prototype selection methods, but in general, the mechanism for selecting prototypes can be broken down into three main categories [16]. The *condensation* methods find the points close to the decision boundaries and remove the internal points. The purpose of this method is to save the classifier's accuracy, but remove the redundancy. *Edition* methods attempt to reduce the noise by removing prototypes near decision boundaries which conflict with its surroundings. Finally, *hybrid* methods combine the two previous methods.

2.2. Dissimilarity representation

Dissimilarity representation is a fast growing field [19]. There have been many recent publications which concern different aspects of dissimilarity based pattern recognition [20–23] as well as extrapolating dissimilarity representation to graphs [24], string based object representation [25], and signature verification [26,27].

There have also been attempts to increase the efficiency of classification and the applications to different patterns. Pekalska et al. [28] proposed using normal density-based classifiers in DSE to overcome the limitations of k-NN methods. Pekalska et al. [29] also analyzed prototype selection techniques, such as mode seeking, k-Centers, an editing and condensing algorithm, linear programming, and feature selection, within DSE. Riesen and Bunke [30] also used prototype clustering in addition to prototype selection techniques on dissimilarity based graph classification. Other trials for prototype selection in DSE include using genetic algorithms [31] and using it for the application of signature verification [32].

3. Dissimilarity space embedding

Instead of characterizing patterns by its features, it is possible to represent patterns as dissimilarities to other patterns [4]. This dissimilarity representation measures the distance between patterns with the consideration that patterns of the same class are more likely to be similar than patters of a different class. Due to variations of patterns, there exist instances where a category of patterns does not share a common feature, but it is still possible to classify the patterns based on the distance to a prototype [33]. By using the distance measure between patterns as features, we can emphasize the fundamental relationship between patterns and de-emphasize the features of the patterns. Dissimilarity representation allows for the embedding of information on how the patterns exist on the system as a whole.

Using this dissimilarity representation, we can take a vector space embedding approach by representing the data as vectors of dissimilarities. An N-dimensional vector space is created from the dissimilarities between each prototype $\mathbf{p}$ in $\mathbf{P}$ and each sample $\mathbf{s}$ in $\mathbf{S}$ based on a distance measure $d(\mathbf{p}, \mathbf{s})$. As illustrated in Fig. 1, the key idea is to select N number prototypes $\mathbf{P}$ and represent any sample as a vector $\mathbf{v}_s$ whose entries are the distances

$$\mathbf{v}_s = (d(\mathbf{p}_1, \mathbf{s}), d(\mathbf{p}_2, \mathbf{s}), \ldots, d(\mathbf{p}_N, \mathbf{s}))^T. \quad (1)$$

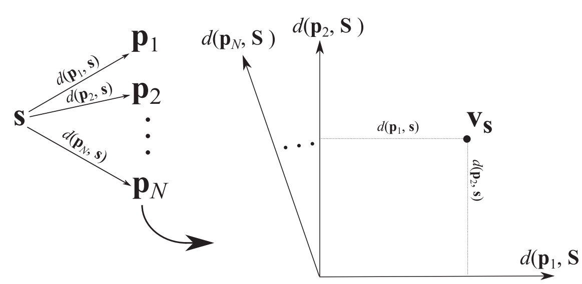

Fig. 1. A visualization of embedding a sample $\mathbf{s}$ into an N-dimensional vector space using the dissimilarities to prototypes $\mathbf{p}_1, \ldots, \mathbf{p}_N$. $\mathbf{v}_s$ is the vector whose entries are the distances to the respective prototypes.