Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 82 (2016) 12 – 19

Procedia Computer Science

Symposium on Data Mining Applications, SDMA2016, 30 March 2016, Riyadh, Saudi Arabia

DTW-Global Constraint Learning using Tabu search algorithm

Bilel Ben Ali[a,*], Youssef Masmoudi[b], Souhail Dhouib[c]

[a] *Faculty of Economics and Management of Sfax, University of Sfax, Tunisia*
[b] *Saudi Electronic University, Riyadh, Saudi Arabia*
[c] *Higher Institute of Industrial Management of Sfax, University of Sfax, Tunisia*

Abstract

Many methods have been proposed to measure the similarity between time series data sets, each with advantages and weaknesses. It is to choose the most appropriate similarity measure depending on the intended application domain and data considered. The performance of machine learning algorithms depends on the metric used to compare two objects. For time series, Dynamic Time Warping (DTW) is the most appropriate distance measure used. Many variants of DTW intended to accelerate the calculation of this distance are proposed. The distance learning is a subject already well studied. Indeed Data Mining tools, such as the algorithm of k-Means clustering, and K-Nearest Neighbor classification, require the use of a similarity/distance measure. This measure must be adapted to the application domain. For this reason, it is important to have and develop effective methods of computation and algorithms that can be applied to a large data set integrating the constraints of the specific field of study. In this paper a new hybrid approach to learn a global constraint of DTW distance is proposed. This approach is based on Large Margin Nearest Neighbors classification and Tabu Search algorithm. Experiments show the effectiveness of this approach to improve time series classification results.

Peer-review under responsibility of the Organizing Committee of SDMA2016

Keywords: Dynamic Time Warping; Tabu Search; Data Mining; KNN classification; Time series; DTW-Global Constraint Learning

1. Introduction

With the availability of enormous time series databases, such as bio-metrics and weather, there has been an explosion of interest in the exploration of these data. Many methods and algorithms have been proposed to classify, index, segment and discriminate time series.

Many classification methods, similarity measures and algorithms have been developed over the years, mostly for survey data. Unfortunately, most of these similarity measures can not be used directly on the time series data sets. New distances and new strategies have been identified, some of which are based on relatively recent tools or results of the analysis of time series: cepstrum coefficients, wavelet transform, hidden Markov models etc. Classification[1,2] is to group objects into classes whose have similar features and content. The objects of the same class are similar and objects from different classes are different. Each classification method is thus based on a ”similarity - dissimilarity”

* Corresponding author. Tel.: +216 97 496 468.
E-mail address: bilel_benali@yahoo.fr

Peer-review under responsibility of the Organizing Committee of SDMA2016
doi:10.1016/j.procs.2016.04.003

measure between objects, a measurement of "similarity - dissimilarity" between classes and aggregation strategy used to build the classes. Many classification methods are available in standard statistical software: partitioning methods (K-means, K-means clustering etc.) Self Organizing Maps and hierarchical methods etc.
Hundreds of distances have been proposed to classify time series, among which Euclidean distance is the most popular. But when it comes to classify time series using Euclidean distance, and any other Minkowski metric, can lead to very intuitive results. In particular, this distance is very sensitive to scale effects, the presence of atypical or missing items and does not take into account possible time lags. One way to solve these problems is to define new distances and similarity measures: Dynamic Time Warping is one of distance measure commonly used for time series data sets. We find that the use of DTW with KNN gave bad results for some instances such as "Swedish Leaf" given by works of Keogh[3]. To improve classification results we will consider to adapt the DTW distance to the studied case using distance learning[4]. The idea is to learn parameters of DTW using a hybridization of the Large Margin Nearest Neighbors and Tabu Search algorithms.
The paper is organized as follows: In the first section it is to present preliminary concepts: Time series and similarity measures, Large Marge Nearest Neighbors (LMNN) classification and Tabu Search (TS) algorithm. The second section is about the proposed approach: using TS algorithm and LMNN classification to learn a DTW Warping Window (DTWWW). A method to condense time series data set used in learning process is presented. The data condensing method minimize the the learning CPU Time. In the fourth section experiments are presented and finally the paper finish by a conclusion and future works.

2. Literature review and related works

2.1. Metric learning review

A lot of work on learning metrics and similarities is about learning the parameters of a Mahalanobis distance. The squared Mahalanobis distance, defined by $D^2_M(x_1, x_2) = (x_1 - x_2)^T M(x_1 - x_2)$, is parameterized by the Positive Semi-Definite(PSD) matrix M. The PSD constraint ensures that D_M is a (pseudo) metric, which allows acceleration of the k-NN classification based on the triangle inequality. Different literature methods differ primarily in the selection of the objective function and the regularization term. For example, in[5], authors forced examples of the same class to be closer than examples of different classes by some margin. In[6] the objective function is related to the error of the k-NN on the training set. Davis and Kulis[7] regulate with the divergence LogDet (which automatically imposes the PSD constraint) while Ying and Huang[8] use the norm (2.1) that promote the learning of a matrix M of low rank. There are also online learning methods, such as POLA[9] and LEGO[10]. The most costly aspect of many of these approaches is the satisfaction of the PSD constraint, although some methods are able to reduce the cost of computing by developing specific solvers.
Some research focuses on learning other types of distances. Qamar[11] optimizes a cosine similarity to treat information retrieval tasks. In the field of image recognition, Frome and Singer[12] learn a local distance for each example, while Chechik and Shalit[13] propose an online learning procedure for bi-linear similarity measure.
The information used in supervised metric learning is of two types: (i) constraints based on pairs of examples: x and y must be similar (or dissimilar), and (ii) the constraints based on examples triples: x must be more similar to y than z. Note that the two types of constraints can be built from labeled data. The objective is to find the metric or similarity that best satisfies these constraints. All methods presented above are generally used in the context of the Nearest Neighbors (NN) (and sometimes clustering). This is due to the fact that constraints based on pairs or triplets are easy to obtain and optimize the sense in the context of the k-NN or clustering algorithms that are based on local neighborhoods.

2.2. Related works

A common way to obtain a family of metrics on a vector space X is to consider the Euclidean distance after the linear transformation $x' = Lx$. These metrics calculate the square distances as given by equation (1).

$$d_L(x_i, x_j) = \|L(x_i - x_j)\|_2^2 \tag{1}$$