UDC 519.23

Learning Deep Euclidean Embeddings for Time Series

A.Kiselev

Moscow Institute of Physics and Technology

We consider the problem of learning a dense euclidean embedding for time series data. Such embedding can be efficiently used as an additional feature vector for forecasting model or for time series classification. We propose a siamese recurrent neural network architectire that is able to produce euclidean embeddings for input data as a side effect of learning given similarity between time series.

Proposed neural network consists of a shared Deep Bidirectional LSTM layer followed by a shared fully-connected layer (fig. 1). The LSTM followed by a dense layer learns a nonlinear mapping from a space of time series into R^n . For a given pair of sequences a predefined distance metric is computed, and neural network is trained to embed time series with respect to that metric. We chose to embed time series with respect to Dynamic Time Warping (DTW), which is a well-known metric that finds optimal alignment between two given sequences. Formally, neural network is trained by minimization of the following loss function:

$$DTW(ts_i, ts_j) - \|f(ts_i) - f(ts_j)\|_2$$

where ts_i, ts_j is a pair of sequences, f - is an embedding function represented by a neural network.

Neural network weights are optimized by gradient descent via RMSprop optimizer. Each batch is a random sample of pairs of time series and respective DTW distances between them.

Embeddings produced by a trained neural network can be used as a feature vectors for time series classification. We tested our approach on three datasets from UCR Time Series Classification Archive, namely TwoPatterns, TwoLeadECG, LargeKitchenAppliances. Three 1-NN (Nearest Neighbours) classifiers are compared on each dataset: 1-NN on raw data with Euclidean metric, 1-NN on neural embeddings with Euclidean metric, 1-NN on raw data with DTW metric. Accuracy scores on test data are reported.

Name	1-NN, Euclidean	1-NN, Embedding	1-NN DT
TwoPatterns	0.910	0.990	0.998
TwoLeadECG	0.747	0.854	0.86
LargeKitchenAppliances	0.493	0.698	0.795

Neural embeddings conserve DTW distances between objects - that is demonstrated by 1-NN classification performance. Dense embedding vectors allow effient usage of advanced classifiers such as Random Forest (RF) or Gradient Boosting Machines (GBM) that are usually unable do deal with raw time series data to further increase classification performance. We were able to reach accuracy scores of **0.99** on each of those datasets by training and fine-tuning GBM (xgboost implementation) on embedding vectors. This is impossible on raw time series data and can be achieved due to dense and heterogenous structure of embeddings.

Evaluation of predictive qualities of such embeddings is in progress. We trained a neural network to approximate DTW distances between 2-hour windows of BTC/USD minutely returns. Such financial data is highly volatile, but still can be efficiently embedded. To demonstrate ability of embeddings to conserve properities of the input data we reconstruct total returns for 2 hour long timeslices from respective 50-dimensional embeddings. It appears that it can be reconstructed with simple Linear Regression model to an coefficient of determination of $R^2=0.98$. That level of accuracy achieved by a primitive model shows that produced representations efficiently conserve some properties of the data, even if it is as volatile as financial time series.

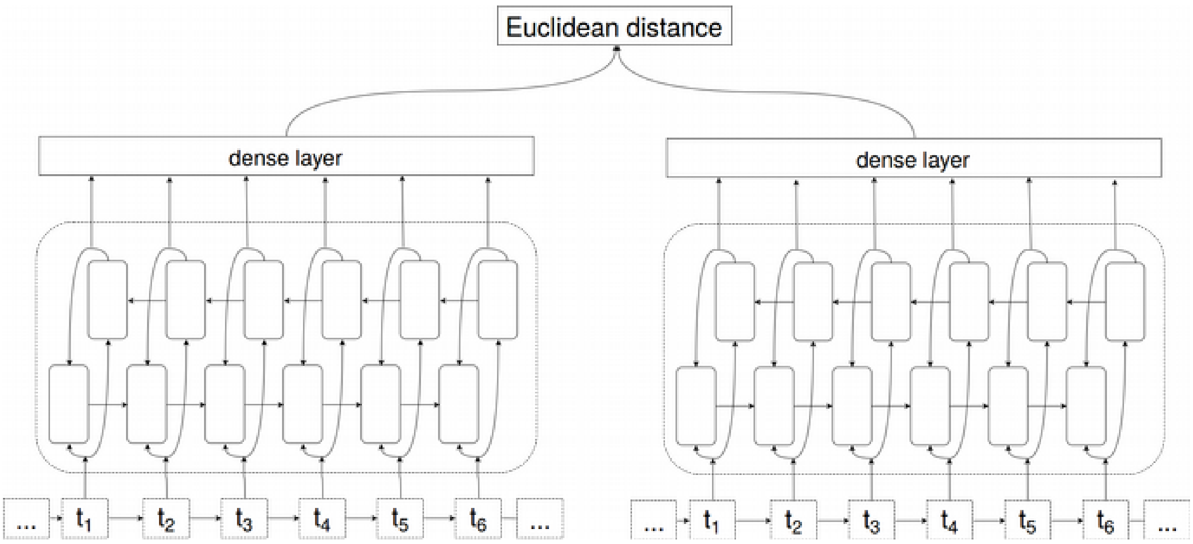

Fig. 1. Siamese BLSTM Architecture

References

1. *Jonas Mueller, Aditya Thyagarajan.* Siamese Recurrent Architectures for Learning Sentence Similarity // Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16) – 2016
2. *Keogh* [*at al.*] The UCR Time Series Classification/Clustering Homepage