Review Article

A STUDY ON PREDICTION BREAK POINTS OF TIME SERIES FORECASTING

PRABAKARAN.N[1]* ,KANNADASAN.R[2]

[1,2] **School of Computer Science and Engineering, Vellore Institute of Technology, Vellore-632014, India,**

***Email: [1]dhoni.praba@gmail.com, [2]desurkannadasanr@gmail.com**

Received: ***Revised and Accepted:***

ABSTRACT

This Study is based on a review conducted on the various models and approaches that aim to predict the time series data using a distance-based prediction approach. The very nature of time series data makes classification a challenging task. Many approaches have been proposed to achieve this one of them being the distance-based approach. 1-NN is the simplest and most widely used method due to its good performance. Over the years new and more complex classifiers have arisen with outperform 1-NN. One of the approaches includes transformation of time series into feature vectors using the distance measure. The feature vectors obtained help to bridge the gap between time series and traditional classifiers. This report includes all of these methods that use a distance-based approach as well as the merits and demerits of each of these methods.

Keywords: Time series, 1-NN, Distance based approach, Machine learning classifiers

DOI: http://dx.doi.org/10.22159/jcr.07.01.01

INTRODUCTION

A data set that represents information over a period of time is termed as a time series dataset. The order of the points in this dataset is very important to understand the behaviour of the dataset. If the order of the points changes, the meaning of the data changes. Such type of data is very different than what the traditional classifiers deal with. Some datasets can be abnormally different from other datasets [1][2]. Identifying such anomalies is becoming increasingly important in every field. Traditional classifiers are not an efficient approach to such type of data. Time series classifiers helps to classify the data based on time period [3]. One of the approaches to classify time series data is using feature extraction. In this method information is extracted from time series data and represented as scalar feature Coordinates. However, this approach can be time consuming and can also cause loss of information.

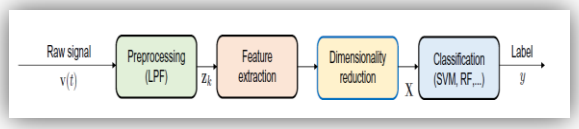

Fig. 1: Operation Flow Of Feature Based Classification Procedures

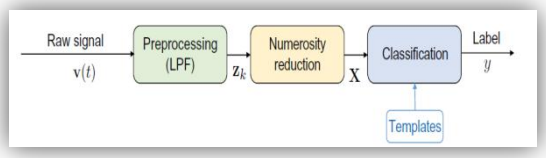

Fig. 2: Operation Flow Of Distance Based Classification Procedures

Another approach is to use a distance based metric classifier. This avoids the feature extraction phase. Generally, the methods are considered into two core categories. 1-NN is the most popular and surprisingly effective choice [4][5]. This can use distances in k-NN classifiers. In another method distances are used to transform the time series data into feature Coordinates.

STUDY REVIEW

A. K-Nearest Neighbour

This approach makes use of the current techniques of time series distances in the k-NN classifier. Mostly, the 1-NN classifier is being used popularly in time series based classification due to its simplicity and good performance [6][7]. The 1-NN classifier predicts the class of a time series using any distance measure such that the class is closest to another class in the training dataset. The k-NN classifier is sensitive to noise which is a commonly found property in the time series data[8].

B. Distance Features

This approach uses the time series distance measures to transform the data into feature Coordinates. This helps in overcoming the particular requirements of classifiers [9]. Some of them include dealing with ordered data and also variable lengths of data. The advantage of this approach is that distance features hold information which is relative to other series data. Whereas the feature-based classifier holds information about the time series in itself[10][11].

C. Global Distance Features

In this method the distance matrix is "i" built by calculation the distance between each sample record. Further, each iteration of row of this distance matrix is represented as a vector that describes the time series data. These Coordinates as given as an input to the classifier.

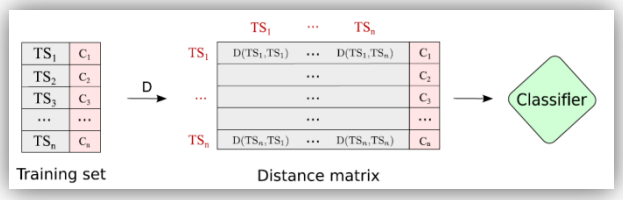

Fig. 3: An Example Of The Global Distance Features Method

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

Inside the classifier the time series in represented using distances to other remaining series in the dataset. A voting schema is used to classify the data using the trained classifiers. Support Vector Machines are the mostly commonly used classifiers for such data [12][3].

D. Local Distance Features

This method uses distances to local patterns in the series instead of the entire series. This method assumes that the discriminatory characteristics are local in nature and such distances are used as features [14][15]. A subsequence of a series that can be identified as a representative of a class are taken into consideration. Such sub sequences are called as shapelets [16]. They are highly interpretable and provide an advantage for the users. This helps the users to understand the meaning of the obtained shapelets.

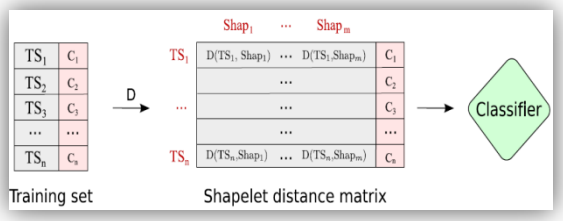

Fig. 4: Typical Example of Local Space Models Using ST

These shapelets are then transformed and fed to the classifier. It can be used in combination with any classifier such as Probability based Naïve Bayes, 1-NN, Bayesian classification Network, Rotation Forest, Random Forest and SVM classifiers [17][18].

E. Embedded Features

This method does not directly work on the distances, instead it makes use of distances to acquire a new depicted representation of the time series valued data [19]. These distances are being utilized to embed the data into a Euclidean distance space while retaining the original distances [20].

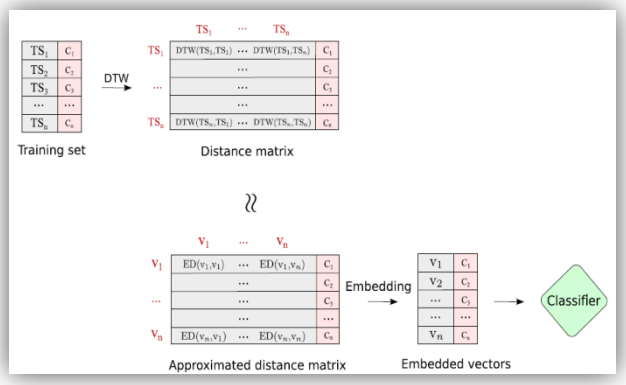

Fig. 5: Example Of The Stages Of Embedded Distance Features Methods

This method gives an advantage in case of time series data since it gives a vector representation of the data [21]. The time series distances are determined by distances between existing vector coordinates. The authors in this review do not use this method to classify time series but for clustering [22]. However, it can be directly applied for classification since it is most likely related to the other distance features methods[23-27].

Table1: Classification and Dependencies

Classifier	Dependencies	Time complexity
K-NN	Size of dataset(m) Distance measure	O(n2m)
Distance based	Distance measure Learning phase	O(n2m2)
Shapelet transformation	Distance measure Learning phase Pre-processing (shapelets:s)	O(nms)

RESULTS AND CONTRIBUTIONS

The computational complexity of each of these methods is an important aspect that needs to be addressed. For any classification algorithm, the time complexity depends heavily on the learning phase of the classifier and the size of the training dataset. However, in case of distance-based classification, the time complexity also depends on the prediction phase and also on the distance measure that has been used. In addition, the cost of the distance measure depends on the length of the time series data. Commonly used distance measures take up too much time while being used in real world applications. In case of the 1-NN classifier the time complexity purely depends on the size of the dataset and the cost of the distance measure. There is no learning phase complexity involved here. The time complexity is O(n2m) (where m is the size of the training dataset).

For the methods that use distance features the computation of learning phase and the cost of distance need to be calculated separately. Each of these have their own computational costs. This method has a complexity of O(n2m2). In the method involving shapelet transformation there is an additional step of pre-processing to obtain the features. This will be added to the computational costs along with the other factors which are included in the previous distance feature method.

FUTURE WORK

In summary, the time series classification methods usually have a quadratic time complexity in both the length of the time series and size of the dataset. If the series is too long or the data is too large the algorithm can take too much time when dealing with real time data. In the case of methods that use distance measures to generate a feature representation of the time series data. These methods bridge the gap between traditional classifiers and time series data.

REFERENCES

1. Abanda, A., Mori, U. & Lozano, J.A. A review on distance-based time series classification. Data Min Knowl Disc 33, 378–412 (2019).
2. Susto, Gian Antonio & Cenedese, Angelo & Terzi, Matteo. (2018). Time-Series Classification Methods: Review and Applications to Power Systems Data. 10.1016/B978-0-12-811968-6.00009-7.
3. R. A. El-Deen Ahmeda, M. E. Shehaba, S. Morsya and N. Mekawiea, Performance Study of Classification Algorithms for Consumer Online Shopping Attitudes and Behavior Using Data Mining. InCommunication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on IEEE, pp. 1344-1349.
4. Bagnall A, Lines J, Bostrom A, Large J, Keogh E (2017) The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. Data Min Knowl Discov 31(3):606–660
5. Berndt D, Clifford J (1994) Using dynamic time warping to find patterns in time series. Workshop Knowl Discovery Databases 398:359–370.

ISSN- 2394-5125 VOL 7, ISSUE 08, 2020

6. Bostrom A, Bagnall A (2014) Binary shapelet transform for multiclass time series classification. Trans Large Scale Data Knowl Centered Syst 8800:24–46
7. Chen Y, Hu B, Keogh E, Batista GEAPA (2013) DTW-D: time series semi-supervised learning from a single example. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining, p 383
8. Chen Z, Zuo W, Hu Q, Lin L (2015b) Kernel sparse representation for time series classification. Inf Sci 292:15–26
9. Corduas M, Piccolo D (2008) Time series clustering and classification by the autoregressive metric. Comput Stat Data Anal 52(4):1860–1872
10. Cuturi M, Vert J (2007) A kernel for time series based on global alignments. IEEE Trans Acoustics Speech Signal Process 1:413–416
11. Fu TC (2011) A review on time series data mining. Eng Appl Artif Intell 24(1):164–181
12. Haasdonk B, Bahlmann C (2004) Learning with distance substitution kernels. In: Joint pattern recognition symposium, pp 220–227
13. He Q, Zhi D, Zhuang F, Shang T, Shi Z (2012) Fast time series classification based on infrequent shapelets. In: Proceedings of the 11th ICMLA international conference on machine learning and applications vol 1, pp 215–219
14. [14] Hills J, Lines J, Baranauskas E, Mapp J, Bagnall A (2014) Classification of time series by shapelet transformation. Data Min Knowl Discovery 28(4):851–881
15. Kaya H, Gündüz-Öüdücü S (2015) A distance based time series classification framework. Inf Syst 51:27–42
16. Keogh E, Kasetty S (2002) On the need for time series data mining benchmarks. In: Proceedings of the 8th ACM SIGKDD international conference on knowledge discovery and data mining, pp 102
17. G.A. Susto, A. Beghi, Dealing with time-series data in predictive maintenance problems,in: 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automa-tion (ETFA), IEEE, 2016, pp. 1–4.
18. M.G. Baydogan, G. Runger, Learning a symbolic representation for multivariate time seriesclassification, Data Min. Knowl. Disc. 29 (2) (2015) 400–422.
19. P.-F. Marteau, Time warp edit distance with stiffness adjustment for time series matching,IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2009) 306–318.
20. L. Ye, E. Keogh, Time series shapelets: a novel technique that allows accurate, interpretableand fast classification, Data Min. Knowl. Disc. 22 (1–2) (2011) 149–182.
21. Z. Chen, W. Zuo, Q. Hu, L. Lin, Kernel sparse representation for time series classification, Inf.Sci. 292 (2015) 15–26.
22. Kannadasan, R., Rajasekaran, K.P., Jaganath, S. and Prabakaran, N., 2019. Performance Analysis of Data Processing Using High Performance Distributed Computer Clusters. Journal of Computational and Theoretical Nanoscience, 16(5-6), pp.2372-2376.
23. S. Velliangiri, P. Karthikeyan & V. Vinoth Kumar (2020) Detection of distributed denial of service attack in cloud computing using the optimization-based deep networks, Journal of Experimental & Theoretical Artificial Intelligence, DOI: 10.1080/0952813X.2020.1744196
24. Praveen Sundar, P.V., Ranjith, D., Vinoth Kumar, V. et al. Low power area efficient adaptive FIR filter for hearing aids using distributed arithmetic architecture. Int J Speech Technol (2020). https://doi.org/10.1007/s10772-020-09686-y,
25. Vinoth Kumar V, Karthikeyan T, Praveen Sundar P V, Magesh G, Balajee J.M. (2020). A Quantum Approach in LiFi Security using Quantum Key Distribution. International Journal of Advanced Science and Technology, 29(6s), 2345-2354.
26. Umamaheswaran, S., Lakshmanan, R., Vinothkumar, V. et al. New and robust composite micro structure descriptor (CMSD) for CBIR. International Journal of Speech Technology (2019), Vol. 23, Issue 2, pp. 243-249.
27. Karthikeyan, T., Sekaran, K., Ranjith, D., Vinoth kumar, V., Balajee, J.M. (2019) "Personalized Content Extraction and Text Classification Using Effective Web Scraping Techniques", International Journal of Web Portals (IJWP), 11(2), pp.41-52