Feature Generation for Time Series

Amiral's Disruptive Technology
versus
the State of the Art

Mazen Alamir

Director of Research at CNRS
Co-Founder of Amiral Technologies

White Paper

June 2018

Amiral Technologies is promoting a disruptive **Feature Generation Tool for time series called DiagSign.** We also promote a package of solutions built around **DiagSign** that solve different industrial predictive maintenance related problems such as the detection of defects, aging and the estimation of the Remaining Useful Life.

Time series are ubiquitous in modern industry thanks to the development of new connected sensors. Analysis of these time series has been subject of active research for decades [2,3] and has been recently considered by [4] as one of the **top 10 must challenging problems** in data mining.

To understand what **DiagSign** brings, let us first recall some facts regarding the use of Machine Learning (**ML**) in Predictive Maintenance (**PM**).

The role of feature generation in deriving predictive maintenance models

The major goal of ML-Based prediction models in PM is to elaborate a prediction (and often a related decision) based on past measurements coming from a set of sensors that are embedded on the equipment. This set of past measurements is the time series we are interested in.

Using the raw version of time series to derive prediction models is generally not efficient and leads to models with poor performances and high level of false alarms. This is because the set of successive values coming from a sensor does not directly reveal variations in the intrinsic properties of the equipment.

To understand this fact, one can think of a time series representing a sinusoidal. As one might know, this signal is characterized by three coefficients, namely, the amplitude, the phase and the period. These three scalars are sufficient to describe any pure sinusoidal signal even if it consists of thousands of points in its raw version. So obviously, elaborating a model with these three parameters is far more efficient, robust (and simple) than elaborating a model with the thousands of values on the time axis of the time series. These three parameters are called the **relevant or discriminant features** that describe the time series.

When one leaves this simple example to the real-life situations where the signals are not so easy to characterize, one can understand that the major challenge is to find **what the relevant/discriminant intrinsic features of a given time series are.** This is obviously a difficult question for which there is no generic answer but our feature generation tool DiagSign is a serious step in that direction.

To understand the novelty of DiagSign, let us take an overlook at existing solutions to the feature generation problem.

These can be classified in the following categories:

First alternative: Structured Feature Generation Approaches

The idea here is to assume that the time series are related to some rule (model) and that this rule is defined through some set of parameters, called hereafter, the model's parameters. Examples of such models are (non-exhaustive list): Auto-Regressive Models [5], Linear Dynamic Systems (LDS) [6], Hidden Markov Model (HMM) [7], conditional Restricted Bolzmann Machine (cRBM) [8], Recurrent Neural Networks [9], Convolutional Neural Networks [10] as well as many other models belonging to the Deep Learning universe.

Note also that any decomposition of the time series on some finite **basis function** belongs to this structured feature generation category in which the parameter of the model corresponds to the vector of coefficients of the time series on the chosen basis. Consequently, all the approaches which are related to the frequency analysis via **Fourier Series** and **Wavelets** can be viewed as instantiation of this structured approach.

Once such a hypothetic model is chosen, the task is to identify the vector of parameters of the model and this vector represents the features that one is looking for. In the case of our simple sinusoidal signal, a linear oscillator model (two-states dynamical system) would have been sufficient and the computation of its three parameters would have achieved the task.

A common fact linking all these approaches is the need to assume some structure!

By this simple fact, one can already loose the war!!! This is because the linearity assumption used in [5, 6], the causality/connection introduced in [7], the structure of the neural network (number of hidden layers, the definition of each layer as well as its firing nonlinear map), the basis of functions used in the projection and finally, the type of Mexican hat used here and there in the Wavelets are all a priori unjustified choices that can make you miss the crucial features. Moreover, except for the neural network related models, this approach leads to a reduction in the size of the information being investigated in the derivation of the prediction model.

Second alternative: Dimensionality reduction via descriptive features

Here, the idea is to reduce the dimension of the data contained in the time series by **selecting some of its properties**. Examples are (again, non-exhaustive list): Piece-wise Linear Approximation (PLA) [11], Piece-wise Aggregate Approximation (PAA) [12], Singular Value Decomposition (SVD) [13], Clipped Data [14], Principal Component Analysis [PCA], Linear Discriminant Analysis [LDA] applied on the raw times series, etc.

In these approaches there is no a priori structure which is assumed behind the scene but a bunch of mechanisms that **condensate the information contained in the time series** according to different processing logics. Each one of these processing logics concentrates on a subset of descriptive properties of the time series (the sequence of slope, the SVD components, the PCA, etc). But they all lead to a number of features that is generally lower than the set of values contained in the time series.

A common feature linking all these approaches is the risk of dropping potentially relevant and discriminant properties that might be originally contained in the time series and which disappear once the latter is processed

Amiral Technologies' solution: DiagSign- Automatic Features Generation Tool

The novelty of **DiagSign**, the Amiral Technologies feature generation tool lies in the fact that it avoids the drawbacks of the previously discussed alternatives, namely:

1. **It is not based on a priori choice of a model structure** that might condemn the success from an early stage of the design process. Rather, a fully non supervised feature extraction algorithm is used based on pure mathematical mapping between the space of time series and an abstract space of lower or higher dimension (it is more frequently higher than lower!).

2. **It is not based on condensing the information** in a low dimensional data; at least not in the early stage of the process. Indeed, **DiagSign** is able to generate thousands of *new features* including a high number of relevant and discriminant features. It is only then that a reduction process (PCA-like or other) can be used to select the relevant features for the task under consideration.

3. When high number of features are targeted, the generation process scales linearly and the **computation of the features is straightforward and cheap** unlike the only alternative option in this case (deep learning) which needs a high number of iterations on a dedicated processor and a dedicated computation machinery (tensor flow or such).

When gathering these advantages, one can come out with a generic (point 1.) and efficient (points 2. and 3.) decision making design process that is unique on the scene.

Finally, these claims are confirmed through industrial case studies involving real-life data.

For more information:

contact@amiraltechnologies.com

www.amiraltechnologies.com

@Amiraltech

Amiral-technologies

References

[1] M. Lankvist, L. Karksson and A. Loufty. /A review of unsupervised feature learning and deep learning for time series modeling/. Pattern Recognition Letters. Vol 42, pp 11-24, 2014.

[2] E. Keogh, S. Kasetty, On the need for time series data mining benchmarks: a survey and empirical demonstration, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 102–111, 2002.

[3] T.G. Dietterich, Machine learning for sequential data: a review, in: Structural, Syntactic, and Statistical Pattern Recognition, Springer-Verlag, pp. 15–30, 2002.

[4] Q. Yang, X. Wu, 10 Challenging problems in data mining research, Int. J. Inf. Technol. Decis. Making 05 (2006) 597–604.

[5] H. Lütkepohl, New Introduction to Multiple Time Series Analysis, Springer-Verlag, 2005.

[6] D. Luenberger, Introduction to Dynamic Systems: Theory, Models, and Applications, Wiley, 1979.

[7] L. Rabiner, B. Juang, An introduction to hidden Markov models, IEEE ASSP, Mag. 3 (1) (1986) 4–16.

[8] I. Sutskever, G. Hinton, Learning multilevel distributed representations for high-dimensional sequences, Technical Report, University of Toronto, 2006.

[9] Y. LeCun, K. Kavukvuoglu, C. Farabet, Convolutional networks and applications in vision, in: Proceedings International Symposium on Circuits and Systems (ISCASa¸r´10), IEEE, 2010.

[10] H. Lee, Y. Largman, P. Pham, A.Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, Adv. Neural Inf. Process. Syst. 22 (2009) 1096–1104.

[11] E. Keogh, M. Pazzani, An enhanced representation of time series which allows fast and accurate classification, clustering and rele- vance feedback, in: Proceedings of the 4th International Conference of Knowledge Discovery and Data Mining, 1998, pp. 239–241.

[12] B.K. Yi, C. Faloutsos, Fast time sequence indexing for arbitrary Lp norms, in: Proceedings of the 26th International Conference on Very Large Data Bases, 2000, pp. 385–394

[13] C. Faloutsos, M. Ranganathan, Y. Manolopoulos, Fast subsequence matching in time-series databases, ACM SIGMOD Rec. 23 (2) (1994) 419–429.

[14] C. Ratanamahatana, E. Keogh, A.J. Bagnall, S. Lonardi, A novel bit level time series representation with implications for similarity search and clustering, in: Proceedings of 9th Pacific-Asian International Conference on Knowledge Discovery and Data Mining (PAKDD'05), 2005, pp. 771–777