Proceedings of the Sixth International Conference on Research in Intelligent and Computing pp. 29–32
DOI: 10.15439/2021R06
ISSN 2300-5963 ACSIS, Vol. 27

Time Series Forecasting with Data Transform and Its Application in Sport

Hao Do*†, Duc Chau*†, Thuc Cai‡, Han Lam‡, Dat Nguyen‡, Tuong Lam‡, Son Tran*†
* University of Science, HCM City, Vietnam
† Vietnam National University, HCM City, Vietnam
‡ OLLI Technology JSC, HCM City, Vietnam
Corresponding author: hao@olli-ai.com

***Abstract*—Forecasting time series data is an exciting challenge. Although being complex, this is a high potential for industrial use. One of the most significant gaps in the forecasting process is the quality of data representation, especially with the time-series data. This paper proposes an effective method using an integral transform that can show hidden information of the time series data. The integral transform exploits data as a composition of many basic functions and then use this set to present the data. Mathematically, this transform converts the data into another space with another feature, showing many properties hidden in the original form. The experimental result demonstrates our suggestion can learn the transformation rules and then can be applied for many applications.**

***Index Terms*—time series forecasting, integral transform, periodic data**

I. Introduction

Many applications such as predicting gold price or predicting stock price can be formed as time series forecasting problems [1]. These applications' data are usually sequences of numbers or vectors corresponding with time. The fluctuated range of data can be vary depending on the kind and properties of data. In a time series forecasting problem, the primary mission is to predict the next value in the series based on the values in the past.

Time series forecasting [2] [3] is very useful in many real-world applications including economic, management, environment, healthcare, stock prediction [4] [5], etc. This problem means constructing a model from available data to predict the next values in the near future. If the data can be analyzed with rules, the predicted values will be approximated with the truth or the value in the real world. These results should be applied to the world to adjust human behaviours and then gain a better state. The results of the forecasting model provide a quantitative method to measure the impact of past and present on the future.

Forecasting is not an easy problem [6]. There are at least two agents that make the predicting process is so complicated. The first one is whether the predicted value can be predicted. It cannot predict that whether an Adidas store will sell off tomorrow. That event mostly depends on the store owner's decision or the store policy. Many forecasting methods focus on finding the rules or the trends in the data to make a guest, so these methods could not work well with the random data or emotion-based data. The second aspect is the ability to collect enough data to build the predicting model. Data should be collected enough in terms of amount and related context. While the amount can be satisfied, the data context is much more challenging to collect and organize. How many things affect the gold price is a tricky question. On the other hand, the way we structure a database with its own contexts is complicated. These two reasons, especially the second one, mainly cause the difficulty of time series forecasting.

We propose a simple method to predict the following values for time series data in this work. Our approach focuses on the comparative periodic data, which is a common data type in real-world applications. With this type, data is followed by some rules or trends. Remarkably, these data are repeated during the time. We exploit these rules with an integral transform [7]. The transform, in nature, is a mathematical way to separate an occasional series into many periodic simple series called basic. The set of basic functions can present all the rules and trends in the data, and then apply these factors to compute the values in the future. With a cheap computing cost, this method can forecast the next value for the nearly repeated data very efficiently.

The remaining of this paper includes four sections. Section two will present the general approach for the problem of time series forecasting. The next section shows our proposed method. This section presents our transform and its application to calculate the next value in the series. Section four is the application of our data transform to sports data. Particularly, we apply the transform method to sports data to exploit its trend. The final section will summarize our work and then make a conclusion.

II. A General Framework for Time Series Forecasting

To deal with a forecasting problem, there are three main stages that we should complete:

- Defining the problem and collecting data: Identifying what the model will predict, data volume, data properties, and how to collect data.
- Building model: Identifying the input and output of the model, what the main processing model should be.
- Predicting the next values: Specifying the way to apply the constructed model to predict the values in the near future.

In the three stages above, the second stage, building model, can be the most important step. The accuracy and performance of the final stage mostly depend on the quality of the built model in this stage. Constructing a right and efficient model will learn and present the data rules very well. This is fundamental for a good prediction.

III. Data transform and its application to data forecasting

In many applications, the data's current form is difficult to analyse or process. We need to transform the data into another form for more informative data. There are many techniques for data processing in general, and integral transform is one of the most popular choices.

Integral transform [7] [8] is a mathematics tool to transform data from this form to another form via an integral operator. Let $f(t)$ and $K(t,u)$ denote the original data, and the kernel, the integral with the kernel $K(\cdot)$ can be defined as:

- Forward transform:

$$Tf(u) = \int_{t_1}^{t_2} f(t)K(t,u)dt \tag{1}$$

- Inverse transform:

$$f(t) = \int_{u_1}^{u_2} Tf(u)K^{-1}(t,u)du \tag{2}$$

Forward transform is used to explore the hidden properties of the data in the other space. The inverse transform is applied to reconstruct the data after being processed in the other space. Generally, we use all these transforms in reality for many purposes.

IV. Transform and forecasting in sport data

We use the integral transform for analyzing the sport data. Particularly, we consider some football teams including Barcelona, Bayern Munich, and Manchester City, which Pep Guardiola coaches. Using the proposed method, we try to predict the result of Pep's team via predicting the average conceded goal per match.

A. Analyzing data

Pep Guardiola is a genius in football, especially with the role of a coach. His teams always have a clear brand identity due to their playing styles. Pep inherits the style from the traditional Netherland style and develops it to reach a new level. Pep and his teams won and lost due to his strategy for every match.

Pep usually spends from three to five years with each team. In the final years in the teams, his achievements are so bad, and then he would be fired. One or two years before, his results were much better and win many champions. On the other hand, the results in the first years are not good compared to the second year.

One of the most aspects, which affects directly to the performance of a team, is the ratio of conceded goals per match. When the teams perform well and win some champions, this ratio is low. If this ratio is high, the teams play too badly and lose too much. The achievements of Pep is present in the table I with the rightmost column. The number for the years from 2008 to 2019 are given due to the statistics while the number for this year (2020-2021) is still a question. The forecasting problem means predicting the value for this year. For more visibly, these values are illustrated in the figure 1.

TABLE I
THE CONCEDED GOAL/MATCH RATE OF THE TEAMS COACHED BY PEP GUARDIOLA

Team	Year	Match	Conceded goal	Conceded goal/match
Barcelona	2008-2009	38	35	0.92
Barcelona	2009-2010	38	24	0.63
Barcelona	2010-2011	38	21	0.55
Barcelona	2011-2012	38	29	0.76
Bayern Munich	2013-2014	34	23	0.68
Bayern Munich	2014-2015	34	18	0.53
Bayern Munich	2015-2016	34	17	0.50
Man. City	2016-2017	38	39	1.03
Man. City	2017-2018	38	27	0.71
Man. City	2018-2019	38	23	0.61
Man. City	2019-2020	38	35	0.92
Man. City	2020-2021	38	?	?

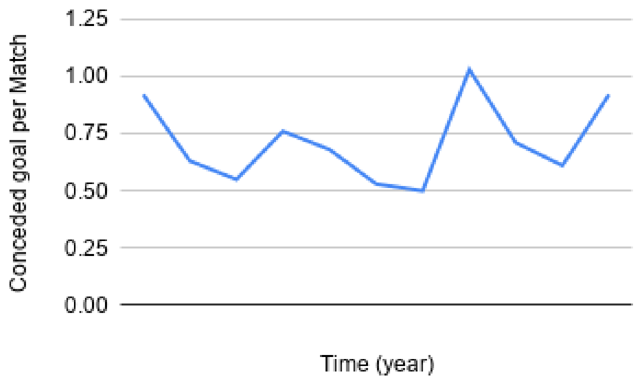

Fig. 1. The average conceded goal per match of Pep's team.

It is easy to observe in the figure 1, the distribution of conceded goal is near a periodic waveform. Pep's achievements can be separated into three-stage corresponding with three teams and three periods in the figure 1. The local bottom values are with the third, seventh, and tenth years. They locate at the two to three years in each team and are the most successful year of these teams. This fact can be interpreted by analyzing Pep's strategy as follow:

- Mostly depending on unique techniques, not strong or power.
- Needing the time for the player to adapt with the strategy.
- First year: Not all player plays well because they need more time.
- The second to the fourth year: The team is formed, they play well and perform full strengths of the strategy.
- The final year: The opponents are familiar with the strategy and counter-work to defeat Pep's team.

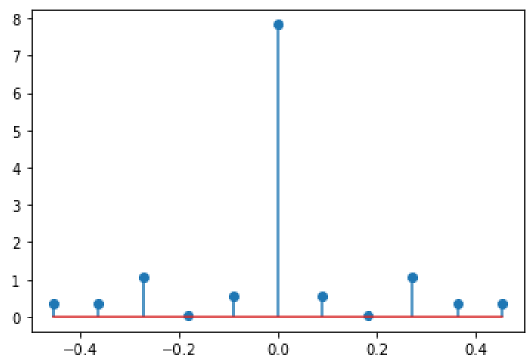

Fig. 2. Frequency representation for Pep's achievement

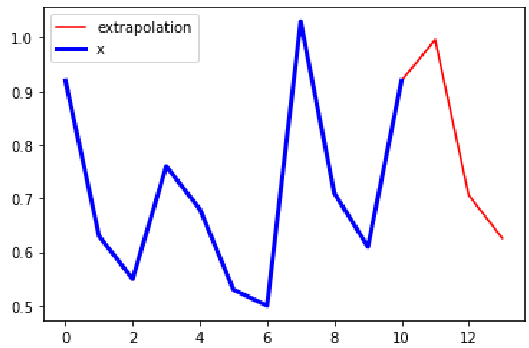

Fig. 3. Predicted value for Pep's team in the next three years

B. Framework for forecasting

As presented in the previous subsection, Pep's teams are usually followed by a fixed strategy and form a fixed achievement curve with three to five years in terms of period. This aspect is very fitted to the strength of Fourier transform. Thus, in this research, we apply this mathematical tool to exploit the data. We apply the below framework for this process:

- Modeling phase: In this phase, we transform the data into the frequency domain via Discrete Fourier Transform [9] [10] with data $f(t)$ and eleven values for ω_k:

$$F(\omega_k) = \frac{1}{2\pi}\sum_{t=1}^{11} f(t)e^{-i\omega_k t} \tag{3}$$

- Forecasting phase: Applying Inverse Discrete Fourier Transform to compute the next value in $f(t)$:

$$f(t) = \frac{2\pi}{11}\sum_{k=1}^{11} F(\omega_k)e^{i\omega_k t} \tag{4}$$

C. Experimental result

In the modelling phase, the frequency coefficients are presented in the figure 2. These elements correspond with the eleven basic waveforms of the data. As can be seen, these values distribute very differently from the original data during the time.

Figure 3 shows the Forecasting phase. After computing the next values in the data series, the results are:

$$f(12) = 0.996 \tag{5}$$

$$f(13) = 0.706 \tag{6}$$

$$f(14) = 0.626 \tag{7}$$

This means that the model predicts Pep's team will not win a lot in 2021 and can lose so much. Let us check the results in reality and then compare with the result showed by the model.

At the end of 2020-2021 season, Pep's team plays 38 matches, and concedes 32 goals. So we have:

$$\hat{f}(12) = 32/38 = 0.842 \tag{8}$$

The season 2021-2022 are continuing (at Dec, 2021) so we can use the current result as a representative. The team has played 20 matches and received 12 conceded goals. So we can estimate:

$$\hat{f}(13) = 12/20 = 0.600 \tag{9}$$

As can be observed, the numbers predicted by our method and their real values are different, but there are two things meaningful in the results. The first is that the differences not too significant. The results can be accepted in a prediction task if the distance from an estimated value to its real value is small enough. The second, and more important, is the trend of sequence value. From $f(12)$ to $f(13)$, in the values estimated by our model, the decreasing rate is:

$$d_r = 1 - \frac{0.996}{0.706} = 0.411 \tag{10}$$

On the other hand, this rate in reality is computed as:

$$\hat{d}_r = 1 - \frac{0.842}{0.600} = 0.403 \tag{11}$$

The trends are decreasing, and the decreasing rates are too similar. This is evidence to prove that our method can exploit many hidden behaviours in the data and predict the values and the trend in the time series data.

V. Conclusion

We have presented an application of integral transform, a simple and effective approach, for time series forecasting problems. This method can exploit many hidden behaviours of the data in a latent space including the rules for change, distribution, and trend. When applied to real data such as average conceded goal, this method performs well with an interesting result. The experiments shows that our proposed method can be a potential solution for time series prediction in real-world applications.

REFERENCES

[1] C. R. Madhuri, M. Chinta and V. V. N. V. P. Kumar, "Stock Market Prediction for Time-series Forecasting using Prophet upon ARIMA." In: 7th International Conference on Smart Structures and Systems (ICSSS), Chennai, India, doi: 10.1109/ICSSS49621.2020.9202042, pp. 1-5, 2020

[2] G. Liu, F. Xiao, C. -T. Lin and Z. Cao, "A Fuzzy Interval Time-Series Energy and Financial Forecasting Model Using Network-Based Multiple Time-Frequency Spaces and the Induced-Ordered Weighted Averaging Aggregation Operation." In: IEEE Transactions on Fuzzy Systems. *28*(11), doi: 10.1109/TFUZZ.2020.2972823, pp. 2677-2690, 2020

[3] S. G. N and G. S. Sheshadri, "Electrical Load Forecasting Using Time Series Analysis." In: 2020 IEEE Bangalore Humanitarian Technology Conference (B-HTC), Vijiyapur, India, doi: 10.1109/B-HTC50970.2020.9297986, pp. 1-6, 2020

[4] Ming-Che Lee, Jia-Wei Chang, Jason C. Hung, and Bae-Ling Chen, "Exploring the Effectiveness of Deep Neural Networks with Technical Analysis Applied to Stock Market Prediction." In: Computer Science and Information Systems, Vol. 18, No. 2, 401–418, 2021, https://doi.org/10.2298/CSIS200301002L

[5] Radojičić, D., Radojičić, N., Kredatus, S., "A multicriteria optimization approach for the stock market feature selection." In: Computer Science and Information Systems, Vol. 18, No. 3, 749–769, 2021, https://doi.org/doi.org/10.2298/CSIS200326044R

[6] N. D. Hieu, N. Cat Ho and V. N. Lan, "An efficient fuzzy time series forecasting model based on quantifying semantics of words." In: 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), Ho Chi Minh, Vietnam, Doi: 10.1109/RIVF48685.2020.9140755, pp. 1-6, 2020

[7] Brian Davies, "Integral Transforms and Their Applications." Springer-Verlag, New York, ISBN 978-1-441-92950-1, 2002

[8] Lokenath Debnath and Dambaru Bhatta, "Integral Transforms and Their Applications." Chapman and Hall/CRC, ISBN ISBN 978-1-584-88575-7, 2016

[9] Smith Steven, "Chapter 8: The Discrete Fourier Transform." The Scientist and Engineer's Guide to Digital Signal Processing (Second ed.), California Technical Publishing, ISBN 978-0-9660176-3-2, 1999

[10] Thomas H. Cormen, Leiserson Charles, Ronald L. Rivest, Clifford Stein, "Chapter 30: Polynomials and the FFT." Introduction to Algorithms (Second ed.), MIT Press and McGraw-Hill, pp. 822–848, ISBN 978-0-262-03293-3, 2001