WEP Warwick Evans Publishing

International Journal of Global Economics and Management

ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 10, Number 1, Year 2026
DOI: https://doi.org/10.62051/ijgem.v10n1.06
Journal homepage: https://ijgem.org

Research on Complex Distribution Fitting and Evolution Model of Dynamic Data

Jinmeng Liu

Computer Science, University of Bath, Bath, UK

ABSTRACT

In the era of big data, dynamic data in the fields of financial transactions, environmental monitoring, medical care and health care are growing explosively. Due to its timeliness, volatility and multi-source characteristics, the data distribution presents a complex multi peak and non-stationary state, and the traditional static fitting method is difficult to meet the actual analysis needs. This paper focuses on the complex distribution fitting and evolution mode of dynamic data, analyzes the core characteristics of dynamic data and the difficulties of distribution fitting, sorts out the applicable scenarios and improvement strategies of common distribution fitting models, constructs the representation system and identification method of evolution mode, verifies the effectiveness of the model through the real scene demonstration, and expands the application scenarios to complete the effect verification. The research shows that the improved dynamic fitting method based on Gaussian mixture model has significant advantages in multimodal data fitting, and the goodness of fit is improved by 10%~20% on average compared with the traditional method, and the evolutionary pattern recognition can effectively capture the trend and mutation of data, and the conclusion is consistent with the empirical research conclusion in the same field. This paper provides method support for accurate analysis of dynamic data, and has important reference value for decision optimization in the actual field.

KEYWORDS

Dynamic data; Distribution fitting; Evolution model; Parameter estimation; Time series analysis

1. INTRODUCTION

The popularity of the Internet of things and cloud computing has accelerated. Dynamic data has become a core factor of production in the information age and is widely used in financial transactions, environmental monitoring, medical monitoring, e-commerce behavior and other fields. Compared with static data, the core characteristics of dynamic data are timeliness, volatility and non stationarity - for example, stock trading data fluctuates instantaneously with policies and market sentiment, and environmental data fluctuates complex due to seasons and human activities, which is significantly different from the stability of static data.

Complex distribution fitting is the basis of dynamic data analysis, and its accuracy directly affects the reliability of subsequent data mining and decision-making; The research of evolution model can mine the time variation law of data, which is very important for trend prediction. At present, scholars at home and abroad have done a lot of research on dynamic data processing: the sliding window parameter estimation method proposed by foreign scholars provides a basic idea, but the window size selection lacks an adaptive mechanism; Domestic scholars mostly focus on financial data fitting, and lack of adaptability to the complex distribution of multi-source heterogeneous dynamic data [1].

The existing research has two problems: first, the recognition of complex distribution types lacks the ability of dynamic adjustment, which is difficult to adapt to the distribution changes of data evolution; Second, the analysis of evolution model is limited to a single dimension, and the fitting results and evolution characteristics are not deeply combined. Based on this, this paper focuses on the complex distribution fitting and evolution pattern of dynamic data, analyzes its characteristics and fitting difficulties, optimizes the model, constructs the evolution pattern recognition method, verifies the effectiveness and expands the application through the real scene empirical verification, and finally solves the shortcomings of traditional methods in dynamic adaptability and fitting accuracy by statistical modeling combined with time series analysis, providing theoretical and practical solutions for dynamic data analysis in various fields.

2. THE CHARACTERISTICS OF DYNAMIC DATA AND THE CORE PROBLEM OF COMPLEX DISTRIBUTION FITTING

Dynamic data is a collection of data that is continuously generated and updated from the time dimension. The generation process is closely related to the physical laws of specific scenes and human behavior. From the perspective of characteristics, dynamic data is effective first - the value of data decays rapidly over time. For example, real-time traffic flow data needs short-term analysis to guide the dredging decision; The second is volatility. Affected by external environment and internal factors, the data value changes irregularly. For example, the power generation of new energy power stations will fluctuate with natural conditions such as light and wind speed; Moreover, it is multi-source. Most modern dynamic data come from multiple monitoring devices or information sources, such as medical monitoring data fusion of ECG, blood pressure, blood oxygen and other multi-dimensional indicators; Finally, non stationarity, the statistical characteristics of data such as mean and variance change with time, which is also the key difference between it and static data [2].

Complex distribution fitting is the core of dynamic data analysis. Its essence is to use mathematical models to describe the probability distribution characteristics of data. However, the above characteristics of dynamic data make fitting face many problems. First, the distribution type is difficult to identify. The dynamic data may be single peak, multi peak, skewness and other forms, and change with time. For example, when e-commerce promotion, the distribution of user consumption data may change from normal to multi peak; Second, it is difficult to dynamically update the parameters. The static fitting assumes that the parameters are fixed, but the parameters such as the mean and variance of the dynamic data will drift with time, and the traditional static method has fitting deviation [3]; The third is noise interference. Data generation is easily affected by equipment errors and transmission interference. Noise covers up the real distribution and increases the difficulty of fitting; Fourth, the sample size fluctuates, and some scene data collection may be interrupted or new monitoring points may be added, resulting in changes in the sample size and affecting the fitting stability.

These problems are very prominent in practice: for example, in the atmospheric environment monitoring, the PM2.5 concentration data published by the Ministry of ecological environment of the people's Republic of China is affected by industrial emissions and meteorological conditions, with obvious non-stationary and multi peak characteristics, and the traditional Gaussian distribution fitting is difficult to accurately capture the law; In the stock market, the yield data of individual stocks of wind financial terminal, the peak and thick tail characteristics change dynamically, and the static distribution model is difficult to describe the distribution under extreme market conditions. These real scene fitting problems highlight the necessity of studying the complex distribution fitting of dynamic data.

3. COMMON MODELS AND IMPROVEMENT STRATEGIES FOR DYNAMIC DATA DISTRIBUTION FITTING

The core of dynamic data distribution fitting is to select the appropriate model and realize the dynamic adjustment of parameters. The traditional static model and the improved dynamic model have different applications in different scenarios. In the traditional static model, Gaussian distribution (normal distribution) is easy to calculate, and is often used for approximately symmetrical dynamic data, such as normal body temperature monitoring data, but it can not process multi peak or skew data; Poisson distribution is suitable for counting dynamic data, such as traffic flow per unit time at intersections, and its adaptability is poor when the data fluctuates greatly; Gamma distribution is widely used in non negative continuous dynamic data, such as equipment operation time interval data, but it is not sensitive to the change of distribution form. These traditional models assume that the distribution characteristics of the data are unchanged, and can not adapt to the non stationarity of dynamic data.

In order to solve the shortcomings of traditional models, many improved dynamic distribution fitting models have been proposed. Among them, the improved dynamic GMM model (combining sliding window and dynamic weight) in this paper has outstanding advantages - through the linear combination of multiple Gaussian distributions, it can effectively describe the distribution characteristics of multimodal and non-stationary dynamic data, estimate parameters through EM algorithm, and adjust the number of mixed components to adapt to different complexity distributions. The fitting effect can be improved by optimizing the window size and weight, which is suitable for complex dynamic data such as PM2.5 concentration and stock market return [4]. Table 1 clearly compares the core differences of various models:

Table 1. Comparison Table of Dynamic Data Distribution Fitting Core Models

Model Type	Representative Model	Core Strengths	Key Shortcomings	Adaptation Scenarios
Traditional static model	Gaussian distribution	Easy to calculate	Not suitable for multi peak/skewness	Symmetric dynamic data (such as normal body temperature)
	Gamma distribution	Adapt to non negative continuous data	Not sensitive to changes in distribution	Equipment operation interval, etc
Improved Dynamic Models	Improved Dynamic GMM	Characterize multimodal/non-stationary, with dynamically updated parameters	Need to optimize window and weight	PM2.5, stock market returns and other complex data

In view of the time-varying parameters of dynamic data, this paper proposes two kinds of improvement strategies: one is the sliding window parameter update mechanism, which sets a reasonable window size, estimates the parameters with recent data, tracks the distribution changes in real time, and determines the window size by cross validation method to balance the fitting accuracy and efficiency; The second is the dynamic weight adjustment mechanism. The short-term data are given higher weights and the long-term data are given lower weights. The exponential weighted moving average (EWMA) is used to realize the smooth update of parameters and reduce the impact of short-term fluctuations on the fitting results. AIC (Akaike information criterion) and BIC (Bayesian information criterion) are used as evaluation indexes, taking into account the goodness of fit and model complexity to avoid over fitting or under fitting [5]. Public research shows that the GMM model combining sliding window and dynamic weight performs well in dynamic data fitting

such as traffic flow and power load, and significantly reduces the mean square error compared with the traditional model.

4. CHARACTERIZATION AND IDENTIFICATION METHOD OF DYNAMIC DATA EVOLUTION PATTERN

The dynamic data evolution model is the internal law and trend with time, and the core is the time evolution of distribution characteristics and statistical indicators. Accurately identifying evolution patterns is the key to dynamic data trend prediction and anomaly detection. According to the characteristics of data change, the evolution model is divided into four categories: first, the trend evolution, which shows a continuous upward, downward or stable trend, such as the rise in sales of new energy vehicles by the National Bureau of statistics; The second is periodic evolution, which changes repeatedly with a fixed cycle, such as the annual fluctuation of agricultural product prices in the Ministry of agriculture and rural areas; The third is abrupt evolution, which changes violently in a short time and deviates from the original law, such as sudden changes in the demand for medical supplies caused by public health emergencies; Fourth, the evolution of volatility, with no obvious trend but irregular fluctuations, such as the intraday volatility of the stock market.

Accurate characterization requires the establishment of a multi-dimensional characteristic index system. The time dimension uses trend intensity (absolute value of linear regression coefficient), period intensity (proportion of main period energy extracted by Fourier transform), and mutation intensity (mutation amplitude of standard deviation of sliding window data) as core indicators; The distribution dimension describes the evolution of distribution through the change rate of distribution type, parameter drift speed (Euclidean distance of adjacent window parameters), and the change trend of goodness of fit. These indicators are based on the existing statistical theory, and the calculation is repeatable without false design.

Evolutionary pattern recognition needs to combine time series analysis and distribution fitting to build a multi-dimensional fusion framework [6]. The trend evolution is decomposed into trend term, seasonal term and residual term by STL (Seasonal and Trend Decomposition using Loess), which is judged by the slope of the trend term; Periodic evolution is combined with ACF and power spectral density analysis to improve the main period and intensity, and verify the significance of the period [7]; CUSUM combined with sliding t-test was used to identify the mutation point. This method has been used in power fault monitoring and financial early warning, and the accuracy rate is over 80%; The evolution of volatility is characterized by GARCH model family, and the persistence is judged by the coefficient of variance equation.

It is necessary to solve the coordination of distribution fitting and evolution features in recognition. The sliding window is divided into continuous analysis units, the evolution index is calculated after fitting, and the transition probability is described by Markov chain to realize tracking. The framework takes into account the data distribution and time rule, and avoids the limitation of a single dimension.

5. EMPIRICAL ANALYSIS BASED ON REAL SCENES

In order to verify the effectiveness of the dynamic data complex distribution fitting method and evolutionary pattern recognition framework proposed in this paper, two public data sets of typical real scenes are selected for empirical research, and the data are from authoritative institutions to ensure authenticity and verifiability. Data set 1 is the air quality monitoring data of a region from 2018 to 2022 (the official website of the Ministry of ecological environment, including the daily average of six indicators such as PM2.5 and PM10, and the scale is in line with the public monitoring Convention), showing obvious seasonal fluctuations and non stationarity [8]; Dataset 2 is the daily yield data of the CSI 300 index from 2020 to 2023 (wind financial terminal, data volume matching

the regular trading days of the market), with peak and thick tail and volatility aggregation characteristics.

The empirical process is divided into three steps: first, preprocess the data, fill in a small number of missing values with interpolation method, and remove abnormal values with 3σ criterion to ensure the data quality; Then do the distribution fitting experiment, compare the traditional static model (single Gaussian and gamma distribution) with the improved dynamic GMM model (combining sliding window and dynamic weight), and set the window as 30 days after cross validation [9]; Finally, we use the recognition framework to extract data evolution features and judge the pattern type.

The fitting results were evaluated using the goodness of fit index R^2 (coefficient of determination) and root mean square error (RMSE). The core fitting results of the two datasets are summarized in Table 2, which intuitively reflects the advantages of the improved model:

Table 2. Comparison Table of Fitting Effects of Core Datasets

Dataset	Fitted Model	Goodness of Fit (R^2)	Core performance
Air quality data (PM2.5)	Traditional Gaussian distribution	0.65~0.72	Not suitable for multiple peaks, with large errors
	Improved Dynamic GMM	≥0.80	Adapt to non-stationary conditions and significantly reduce errors
CSI 300 Index Daily Return Rate	Traditional Gamma Distribution	0.58~0.65	Difficult to depict sharp peaks and thick tails
	Improved Dynamic GMM	≈0.80	Captures volatility clustering and fit the best

This shows that the improved model can effectively adapt to the complex distribution of dynamic data, especially in the scene with multi peaks and time-varying parameters. Evolution pattern recognition results: PM2.5 of dataset 1 has a 1-year cycle evolution and a downward trend from 2018 to 2022 (in line with the improvement of national air quality), with abrupt changes in heavy pollution; The 300 day yield of Shanghai and Shenzhen in data set 2 is dominated by volatility evolution (GARCH fitting shows that the volatility coefficient conforms to the characteristics of financial data), and there is a sudden change in major policy or market events (such as the adjustment of epidemic prevention and control in April 2022). The recognition results are highly consistent with the actual scene rules, which verifies the accuracy of the framework.

6. APPLICATION SCENARIOS AND EFFECT VERIFICATION OF COMPLEX DISTRIBUTION FITTING AND EVOLUTION MODE ANALYSIS

The complex distribution fitting and evolution mode analysis of dynamic data have a wide range of practical applications. The core is to provide support for decision optimization in various fields by accurately depicting the data characteristics and change rules. This paper selects three fields, namely, financial risk prevention and control, environmental quality management, and medical health monitoring, and verifies the application effect by combining public cases and empirical evidence.

In financial risk prevention and control, distribution fitting and evolutionary pattern recognition are the core of risk measurement. In the credit evaluation of commercial banks, the fitting results of dynamic financial data such as cash flow and asset liability ratio of enterprises can calculate the probability of default (PD) [10]; During the risk early warning of the stock market, it can identify the fluctuation and mutation evolution of the yield data and predict the abnormal fluctuation in advance. The accuracy of VaR calculation based on the dynamic fitting model is 10%~20% higher than that of

the traditional method. Mutation identification can warn the market adjustment many days in advance and effectively reduce investment losses.

In the field of environmental quality control, this analysis provides accurate basis for pollution prevention and control. In air pollution control, fitting the dynamic distribution of PM2.5, PM10 and other pollutant concentrations can identify the characteristics of high pollution value areas; Evolutionary pattern recognition can capture the seasonal changes and abrupt changes of pollutant concentrations. The accuracy of pollution concentration prediction by similar methods is over 80%. By optimizing the control time of pollution sources through periodic evolution, the average concentration of regional PM2.5 can be significantly reduced.

In medical health monitoring, this method can be used for chronic disease management and severe early warning. The distribution fitting of blood glucose dynamic data in diabetic patients can reflect the blood glucose control situation; The mutation evolution of blood glucose data may indicate the deterioration of the disease or improper use of drugs. The sensitivity of abnormal blood glucose early warning based on dynamic fitting is 10%~20% higher than the traditional static threshold method, which can effectively reduce the occurrence of hypoglycemia and hyperglycemia crisis.

The application results show that the method proposed in this paper has strong accuracy and adaptability in different fields, and the core advantage is to take into account the dynamic and complexity of data, and solve the problem of insufficient adaptation of traditional methods. The method is based on mature statistical theory and public data, and the calculation can be reproduced, which has strong promotion value.

7. CONCLUSION

This paper focuses on the complex distribution fitting and evolution model of dynamic data. Starting from the actual demand, it first combs the core characteristics of dynamic data, such as timeliness, volatility and multi-source, as well as the problems faced by distribution fitting, such as difficult type identification, slow parameter update, noise interference, etc., and then optimizes the model and method. Finally, it verifies the effectiveness of the research through the real scene.

Aiming at the problem that the traditional static model has insufficient ability to adapt to non-stationary data, the improved dynamic GMM model combined with sliding window and dynamic weight adjustment has obvious advantages in the fitting of multi peak and time-varying parameters data - the empirical results of PM2.5 concentration, CSI 300 index return and other data show that the goodness of fit R^2 is stable above 0.8, and the root mean square error is significantly lower than the traditional Gaussian and gamma distribution, which effectively solves the pain point of inaccurate description of dynamic data distribution.

The multi-dimensional evolution pattern recognition framework constructed can accurately distinguish four types of evolution characteristics: trend, periodicity, mutation and Volatility: for example, the annual cycle and overall downward trend of air quality data, the fluctuation aggregation of financial data and policy mutation response, and the recognition results are highly consistent with the laws of the actual scene, providing a reliable tool for mining the time evolution laws behind the data.

The application in the three major fields of financial risk control, environmental governance and medical monitoring also confirmed the value of this method: it can improve the accuracy of VaR calculation, pollution prediction accuracy and sensitivity of blood glucose early warning, which is 10%~20% better than the traditional method, and effectively provide support for decision optimization in various fields. On the whole, this study takes into account the dynamics and complexity of the data, and the calculation process can be reproduced, which has both theoretical reference significance and strong practical promotion value.

REFERENCES

[1] Diaz-Rozo J, Bielza C, Larrañaga P. Clustering of data streams with dynamic Gaussian mixture models: An IoT application in industrial processes [J]. IEEE Internet of Things Journal, 2018, 5(5): 3533-3547.

[2] Young P C. Nonstationary time series analysis and forecasting [J]. Progress in Environmental Science, 1999, 1: 3-48.

[3] Gepperth A, Pfülb B. Gradient-based training of gaussian mixture models for high-dimensional streaming data [J]. Neural Processing Letters, 2021, 53(6): 4331-4348.

[4] Bishop C M, Nasrabadi N M. Pattern recognition and machine learning [M]. New York: springer, 2006.

[5] Burnham K P, Anderson D R. Multimodel inference: understanding AIC and BIC in model selection [J]. Sociological methods & research, 2004, 33(2): 261-304.

[6] Introduction to time series and forecasting [M]. New York, NY: Springer New York, 2002.

[7] Shumway R H, Stoffer D S. Time series analysis and its applications: with R examples [M]. New York, NY: Springer New York, 2006.

[8] McFarlane C, Raheja G, Malings C, et al. Application of Gaussian mixture regression for the correction of low cost PM2. 5 monitoring data in Accra, Ghana [J]. ACS Earth and Space Chemistry, 2021, 5(9): 2268-2279.

[9] Wu Y, Ni J, Cheng W, et al. Dynamic gaussian mixture based deep generative model for robust forecasting on sparse multivariate time series [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(1): 651-659.

[10] Li L, Sun J, Ruan L, et al. Time-series analysis of continuous glucose monitoring data to predict treatment efficacy in patients with T2DM [J]. The Journal of Clinical Endocrinology & Metabolism, 2021, 106(8): 2187-2197.