Research and Application of Mathematical Model under Big Data

Xizhe Long
College of Science, Tianjin University of Technology, Tianjin, China

Abstract: As a bridge connecting abstract theory and real world application, mathematical model has become an effective tool to solve all kinds of problems. It embodies the essence of mathematics and expands mathematical theory while realizing interdisciplinary interaction. It has promoted the development of science and technology. At present, the era of big data has arrived, and its massive, high-dimensional and heterogeneous data puts forward higher requirements for the establishment of mathematical models. Therefore, standing on the node of the past and the future, this paper analyzes and summarizes the application of mathematical models in various aspects in the current era of big data from the three aspects of financial market risk measurement, big data medical analysis and agriculture, and finally discusses the shortcomings of the current development and makes prospects for the future.

Keywords: Mathematical Model; Risk Measurement; Big Data Medical Analytics; Agriculture

1. Introduction

The advent of the era of big data has brought unprecedented opportunities and challenges for scientific research and social governance. Therefore, the construction, optimization and application of mathematical models are well suited to the problems faced by complex systems. At present, the degree of application of mathematical models at home and abroad is different, and a pattern of multidisciplinary integration has been formed in the world, and important breakthroughs have been made in basic theory and technical application. For example, Kiran Chaudhary, Mansaf Alam, Mabrook S. Al-Rakhami and Abdu Gumaei collect data from various platforms to better predict consumer behavior from the perspective of mathematical modeling through predictive big data analysis. Carolyn McGregor, Mikael Eklund, Khalil El-Khatib, Anirudh Thommandram of the University of Ontario Institute of Technology proposed an expanded version of the Artemis system for analysis and deployed in local hospitals. And obtained important performance metrics in different configurations. However, there are also some problems in foreign research, such as insufficient data privacy protection and poor interpretability of complex models, which require a balance between theoretical innovation and technical ethics. Domestic research has developed rapidly in recent years, driven by application and supported by policies. In the field of medical and health care, Chinese scholars Ma Xiaofan, Gu Shuaishuai, Jin Yansheng, Eman Ghonaem, and Ahmed Mohamed Hamad Arbab, have combined gray prediction mathematical models with clinical big data to build infection risk prediction systems (such as pathogen analysis in patients with lupus nephritis). Providing localized solutions for precision medicine. Chen Kunding, in the realm of big data, introduced a classification system for big data, utilizing the scoring approach of classification mining mathematical models to enhance data mining's efficiency and precision. On the other hand, ChenErfang improved the linked data mining model on the basis of the partial differential classification mathematical model, which provides new ideas for the classification and mining of big data. In the field of smart cities, a team from Tsinghua University used spatio-temporal data models to optimize urban traffic scheduling, significantly reducing the congestion index. In finance, Annie Tu points out how future mathematical theories can be applied to financial risk measurement models, based on traditional mathematical models of risk measurement. However, domestic research still faces bottlenecks such as dependence on imported core algorithms and lack of original theories, especially in the efficient modeling and real-time computing technology of ultra-large scale data.

2. The Application of Mathematical Models in Practice

2.1 Financial Market: Risk Measurement Perspective

The advent of the era of big data has injected new vitality into the financial market, giving birth to brand new financial models such as Internet finance. At the same time, traditional risk measurement models also need to keep pace with The Times and constantly improve. Therefore, by introducing mathematical models and combining the historical data obtained by big data, we can obtain more accurate mathematical models to better measure the risks of various assets [1].

There are three kinds of traditional risk measurement models: first, mean-variance model

Objective function:

$$\sum_{i}^{n} xiri \tag{1}$$

$$min\, \sigma p^2 = \sum_{i=1}^{n} \sum_{i=1}^{n} COVij\,(xi,xj) \tag{2}$$

Condition 1= (Short selling allowed) or $1 = \sum xi$, $\sum xi \geq 0$ (short selling not allowed), where rp , σp^2 , xi and ri respectively are the expected return rate of the portfolio, portfolio risk, the weight of the i asset in the portfolio, and the expected return rate of the i asset), The second is the VaR model(where P, $P\Delta P$, VaR and a respectively represent the probability that the asset value loss is less than the maximum possible loss, the value loss of z assets in a certain holding period, the maximum possible loss, and the given confidence level), and the third is the consistent risk measure (for a risk measurement method, any two portfolios with a value of, meet the following properties called the consistent risk measure:

1. Monotonicity: If $\omega 1 \leq \omega 2$.then $\rho(\omega 1) \geq \rho(\omega 2)$
2. Homogeneity: $\rho\,(h\omega) = h\rho(\omega), h > 0.3$
3. Displacement invariance: $\rho(\omega + k) = \rho(\omega) - k$ k for the combination to increase the amount of cash
4. Additivity: $\rho(\omega 1 + \omega 2) \leq \rho(\omega 1) + \rho(\omega 2)$

We know that risks are complex and uncertain, so the risks of financial assets in the era of big data are different from those in the past, which has high research value. The Internet is transmitting massive data information all the time, and the massive information flow has a great impact on maintaining the stability of the financial market, and it is easy to produce the Domiro domino effect of "affecting the whole body from one start". Therefore, technicians need to capture valuable data and get the characteristics and rules of risks by processing the data. At present, the main means of measuring financial risk is the connection and transformation of the characteristics of data structure panel data model and function. Therefore, we need to put forward a risk measurement theory of applying mathematical models to financial markets in the era of big data: Copola theory [2]-the main content is the principle of the risk measurement of the function [3] : A multivariate joint distribution function can be decomposed into a number of marginal distribution functions, and then a Copula function is used to reduce the cost of establishing a functional model. In the process of Copula function application, the construction of multivariate functions in mathematical theory is used. And the nonlinear and asymmetric relationship between the variables, the edge distribution and correlation are separated, and the correlation relationship of the tail distribution in mathematical statistics can be accurately described. This theory makes the setting of mathematical models more consistent with the law of data. At present, this theory is widely used in the field of financial market, such as portfolio risk calculation, premium analysis of financial law fluctuations and so on.

2.2 Application of Mathematical Model in Agricultural Big Data

As we all know, China has a very long history of farming, and agricultural big data is continuously accumulated in all aspects of agricultural development, including but not limited to the agricultural planting experience of "24 Solar Terms Song". Agricultural big data refers to the collection of all data generated in agricultural production, operation and product sales, and with the sowing, growth, harvest, processing and sale of crops, big data will be continuously updated and accumulated. At present, many problems faced by modern agriculture can be solved through big data application research. However, considering the massive and complex characteristics of big data, it is necessary to embed mathematical modeling to solve the "pain points" of big data.

In modern agricultural production, the

popularization of mechanization has made sowing and harvesting easy, the use of fertilizer can promote the mass production of crops, and the application of biotechnology has made the control of pests and diseases continue to move forward, but agricultural production has always been unable to get rid of the uncontrollable factor of meteorological disasters. Therefore, the application of mathematical model technology provides a new idea for the prediction of meteorological disasters. In the past, due to the large and complicated volume of meteorological data, the efficiency of data processing and classification is low, so it has not been popularized. To solve this problem, the application of K-nearest neighbor combination classifier and distributed parallel processing method has made a great contribution to the establishment of agricultural meteorological disaster classification model. By the parallel K nearest neighbor combination classifier, professionals can establish a practice path of "according to the meteorological disaster grade index-put forward the meteorological disaster risk index formula-complete the statistical classification of agricultural meteorological disaster risk index-realize the disaster risk assessment of crops". With relatively reliable meteorological disaster assessment information, people can predict and protect the meteorological disasters in the future period of time, on this basis, rational arrangement of agricultural production. This not only can avoid unnecessary property losses, but also has a certain positive significance for the supply and demand balance of the entire agricultural product market. Modern agricultural practice shows that the application of mathematical model technology to the data mining and analysis of agricultural big data is helpful to sum up experience and lessons for the "difficult and complicated diseases" in agricultural development, and make law summary and trend prediction on this basis, so as to provide a reference basis for agricultural decision-making from the country down to farmers. This is not only helpful for individual farmers to reduce losses and improve income, but also has great significance for the country to play the advantages of agricultural big data resources, improve agricultural production and management efficiency, and achieve high-quality agricultural development.

At present, China's agriculture is in acritical period of transformation from traditional agriculture to modern agriculture. The development of big data technology is in the ascendant, and mathematical model technology has gradually attracted the attention and favor of more and more professionals. We have reason to believe that the coupling of mathematical model technology and agricultural big data will burst out more sparks and vitality in the future.

2.3 The Application of Mathematical Model in Big Data Medical Treatment-Starting from Two Cases

2.3.1 Neonatal intensive care: introduction to artemis cloud framework and analytical modeling

The data generated by medical devices in neonatal intensive care is a big data problem. Vital organ monitoring, ventilation support, and nutrient or drug titration via smart infusion pumps all generate large amounts of data at high frequencies. However, traditional recording protocols, whether paper or electronic health records, can only store one indicative heart rate per hour for a sustained period of time. High-speed physiological data is an underexploited resource in the medical field. Through policies that promote the systematic adoption of big data approaches for neonatal intensive care, there are significant opportunities in the healthcare sector to make major medical discoveries, improve the quality of care, and thereby improve the efficiency of care, while reducing mortality and disability. The effective use of big data in neonatal intensive care has great potential to support a new wave of clinical discovery and create new online health analytics for disease prevention or earlier detection of symptoms.

Through our research, we have proposed the Artemis Platform [4,5], a big data approach to online health analysis for high-speed physiological data, which can be integrated into other clinical data as needed. The Artemis platform enables concurrent processing of multiple patients, multiple data streams and multiple diagnoses through time analysis to support real-time clinical decision making and clinical research. We have also further proposed the Artemis Cloud as away to deliver Artemis services through the cloud. We designed an extended Artemis cloud platform to serve multiple healthcare organizations, but we needed an analytical model to plan the usage capacity of the platform. The analytical model in the context of cloud-based big data solutions is currently an under-researched area, especially for its application in the healthcare sector. Next, the analytical modeling of Artemis is presented: The Artemis cloud platform employs a hybrid queuing model (M/HE/m/m),(M stands for Markov) [6] where patient arrival intervals follow an exponential distribution, while hospitalization durations adhere to a hyper-exponential distribution (HE). By implementing a four-stage hyper-exponential distribution to simulate variations in length of stay across different patient types, the system dynamically computes bed occupancy rates and blocking probabilities, thereby optimizing resource allocation strategies. We use this case for numerical validation. Therefore, the mean length of hospital stay is:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)\,dx = \sum_{i=1}^{n} pi \int_0^{\infty} x\lambda i\, e^{-\lambda} dx = \sum_{i=1}^{n} {}^{pi}/_{\lambda i} \quad (3)$$

Where, $pi$and $\frac{1}{\lambda i}$are respectively the probability of becoming a Class i patient and the corresponding average length of stay in the NICU. Therefore, the type of queuing system that we need to solve and get

the performance indicator is M/HE/m/m.

Next, we will define the parameters of this queuing model and explain how to estimate certain parameters using healthcare data.

(1) Parameter Definitions

1.Arrival Rate Parameter **(λ)**: Patient arrival intervals follow an exponential distribution with parameter (λ), representing the expected number of patient arrivals per unit time.

2.Hyper-Exponential Distribution Parameters **($\lambda i,\mu i$)**: Hospitalization duration is modeled as a mixture of (k) independent exponential distributions. Here αi denotes the proportion of Class i patients (satisfying $\sum_{i=1}^{k} \alpha i = 1$), and μi represents the discharge rate for Class i patients (mean hospitalization duration = $\frac{1}{\mu i}$).

3.Number of Service Stations (m): The system's parallel service capacity, corresponding to the total number of available NICU beds.

4.System Capacity Limit (m): The maximum patient capacity, typically equal to the number of beds, i.e., excess requests are rejected.

(2)Parameter Estimation Methods

1.Estimation of Arrival Rate (λ**)**: Calculate the mean number of patient arrivals per unit time (e.g., hourly/daily) using historical admission records. Validate the exponential distribution assumption via the Kolmogorov-Smirnov test.

2.Estimation of Hyper-Exponential Parameters ($\alpha i,\mu i$): Patient Classification: Categorize patients into mutually exclusive classes based on clinical features (e.g., gestational age at birth, type of infection, respiratory support level). Classification criteria should be defined collaboratively with medical experts.Parameter Fitting: For each patient class, iteratively optimize the mixture distribution parameters using the Expectation-Maximization (EM) algorithm to maximize the likelihood function:

$\tau(\alpha i,\mu i)=\prod_{j=1}^{N} \sum_{i=1}^{k} \alpha i\, \mu i e^{-\mu i t j}$ where tj is the hospitalization duration of the j-th patient, and N is the sample size.

3.Determination of Bed Capacity (m): Directly adopt the hospital's actual NICU configuration. For example, if a hospital has 20 NICU beds, set (m = 20).

Multiple queuing systems that characterize the non-exponential distribution of service time are not entirely easy to deal with; However, since the M/HE/m/m queue system has no additional capacity beyond the service facilities, it works exactly the same way as the M/M/ M queue system. [7] Fortunately, the steady-state probability of such a queuing system is known and can be obtained in closed form:

$$pk=\frac{\frac{\rho^k}{k!}}{\sum_{n=0}^{m} \frac{\rho^n}{n!}} \quad (4)$$

Where, m is the number of beds. In this model, blocking is when, as described above, an admission request to the Hospital for Sick children results in the patient being referred to one of the other two hospitals with a tertiary NICU, and possibly to a fourth hospital if that network is full. The blocking probability can be obtained by the following formula:

$$pk=\frac{\frac{\rho^m}{m!}}{\sum_{n=0}^{m} \frac{\rho^n}{n!}} \quad (5)$$

The probability generating function (PGF) is:

$$P(z)=\sum_{k=0}^{m} pkz^k \quad (6)$$

The effective patient arrival rate (i.e. the rate of patients able to enter the NICU) is:

$$\lambda e=\lambda(1-Pb) \quad (7)$$

Now, we can derive the required performance indicators. The average number of patients in a NICU is the first derivative of (P(z)) at (z = 1) :

$$\bar{p}=p'(z)z=1 \quad (8)$$

According to Little's Rule [8], the average length of hospital stay for patients is:

$$\bar{t}=\frac{p}{\lambda e} \tag{9}$$

2.3.2 Impact of E. coli infection on lupus nephritis sufferers: using a gray prediction mathematical model

Systemic lupus erythematosus (SLE), a chronic autoimmune disorder characterized by multi-organ involvement [8], arises from complex interactions between genetic predisposition, hormonal influences, environmental triggers, infectious agents, pharmacological factors, and dysregulated immune responses. This condition manifests as pathological immune activity against healthy tissues. Renal complications of SLE, clinically identified as lupus nephritis (LN), represent a severe progression marked by bilateral immune-mediated renal injury across diverse histopathological classifications, often presenting with significant functional impairment [9]. The insidious onset and nonspecific clinical manifestations frequently result in delayed detection and therapeutic intervention, thereby compromising treatment efficacy and patient prognosis while increasing mortality risks. Contemporary advancements in digital health technologies and connected medical systems empower healthcare institutions to leverage comprehensive clinical datasets for enhanced diagnostic precision and epidemiological forecasting. Within predictive analytics, grey system theory offers a robust framework for modeling complex systems with incomplete information. This methodology employs differential equation-based models to quantify developmental disparities among system variables through sequence operator analysis. By transforming raw observational data into structured time-series patterns and establishing corresponding dynamic equations, it enables probabilistic forecasting of system trajectories.Emerging evidence highlights the growing application of grey modeling in healthcare management and clinical prediction sciences. The operational principles of this computational framework can be summarized through three key phases: data preprocessing through accumulation generation, system behavior characterization via grey relational analysis, and predictive modeling using differential equation solutions. This analytical approach demonstrates particular efficacy in scenarios with limited data availability and inherent system uncertainties. The core of grey prediction theory lies in constructing the GM(1,N) model, which integrates dynamic relationships among N variables through first-order differential equations. Specifically, its modeling process consists of two steps: Accumulated Generating Operation (AGO) and solving grey differential equations. First, the original sequence undergoes accumulation processing to weaken randomness. Subsequently, parameters are solved based on the whitening equation. Finally, predicted values are restored through inverse accumulated generation.

Suppose the initial sequence is:

$$S^{(0)}(t)=\{S^{(0)}(1),S^{(0)}(2),S^{(0)}(3),\ldots S^{(0)}(n)\} \tag{10}$$

Then add the pairs to get equation (11) : $S^{(0)}$

$$S^{(1)}(x)=\{S^{(1)}(1),S^{(1)}(2),S^{(1)}(3),\ldots S^{(1)}(n)\} \tag{11}$$

Where

$$\sum_{y=1}^{x} S^{(0)}(y) \tag{12}$$

x=1,2,3...n

$T^{(1)}(x)$ Yes immediate mean generates a sequence whose sequence is: $S^{(1)}(K)T^{(1)}(x)=\{T^{(1)}(1),T^{(2)}(2),T^{(3)}(3)\ldots T^{(1)}(n)\}$(13)

Among x=2,3...n

Equations containing grey derivatives or grey differential equations are called grey differential equation and their differential equations are:

$$dS^{(1)}/_{dt}+iS^{(1)}=j \tag{14}$$

When i and j are undefined parameters, their values can be determined by least square method and vector parameters can be expressed by equations

$$\bar{\Lambda}=(j)^T=(B^TB)^{-1}B^TY \tag{15}$$

This paper proposes a differential equation accumulation forecasting model to improve prediction accuracy and effectiveness.

$$S^{(0)}(x+1)=\{S^{(0)}(1)-j/i\}e^{-ix}+j/i \tag{16}$$

By using this model to diagnose the infection of LN patients, generally a higher diagnostic rate can be obtained.

3. Conclusion

This paper mainly analyzes the application scenarios of mathematical models under big data from three aspects: financial market, agricultural big data and big data medical treatment. First of all, the financial market achieves accurate asset risk measurement by combining the historical data of big data, and puts forward the Copola theory that the setting of mathematical model conforms to the law of data. Mathematical modeling can solve the "pain point" of agricultural big data, and establish the agricultural meteorological disaster classification model by using K-nearest neighbor combination classifier and distributed parallel processing method, which provides a new way for meteorological disaster prediction, helps to rationally arrange agricultural production, avoid property losses and maintain the balance between supply and demand of agricultural products market, and provides a reference for agricultural decision-making. In the aspect of big data medical treatment, Artemis platform and Artemis cloud are proposed, and various indicators obtained through this way are helpful to real-time clinical decision-making and research. The gray prediction mathematical model can identify the difference in the development trend of system factors, and predict the development trend of things through the processing and modeling of the original data, which can be used to diagnose the infection of LN patients to obtain a higher diagnosis rate. The above indicates the innovativeness of the research method. The innovation also lies in the realization of interdisciplinary application expansion, giving full play to the role of mathematical model as a bridge connecting abstract theory and real world application, and promoting the development of various fields.

Of course, there are still shortcomings in this paper, for example, as data scales grow exponentially (e.g., petabyte-level time-series data integration across regional healthcare systems), the proposed models face the following critical bottlenecks: High-Dimensional State Space Explosion: In multi-institution collaboration scenarios, the state space dimensionality of the hybrid queuing model (M/HE/m/m) grows combinatorially with the number of hospitals, leading to computational complexity of $O(m^{2k})$ for traditional dynamic programming algorithms, severely limiting real-time decision-making capabilities. Heterogeneous Data Synchronization Delays: In distributed deployments, aligning physiological signal data with electronic health records across networks introduces 10-15% clock drift errors even in 5G edge computing environments, degrading model convergence stability. In addition, potential solutions for distributed computing should be adapted:1. Hierarchical Distributed Architecture: Implement edge-center computational pipelines using stream processing frameworks, where edge nodes handle lightweight preprocessing, and central clusters execute high-throughput parallel simulations.2.Parameter Parallelism and Model Sharding: Leverage distributed training frameworks to shard parameters and reduce communication overhead via asynchronous parameter servers.3.Approximate Computing and Lightweight Modeling: Integrate techniques like Monte Carlo sampling and tensor decomposition to compress computational workloads while balancing accuracy and efficiency. Heterogeneous Resource Scheduling Dynamically allocate CPU/GPU resources using container orchestration platforms, combined with acceleration libraries for compute-intensive tasks. For the future research direction, I think we should focus on the following points: 1. Strengthening data security and privacy protection. Improve the scope of application of mathematical models. The breakthrough of real-time calculation and efficient modeling technology.

References

[1] He Hongqing. Discussion on the application of Mathematical Model in financial Market [J]. Science and Technology Economic Market,2009(01).

[2]Zhao Liqin. Research on Financial risk Measurement based on Copula function [D]. Xiamen University,2009.

[3] Ma Wei. A method of financial risk measurement under the condition of Big Data [J]. Statistics and Decision,2015(02).

[4] M. Blount, M. R. Ebling, J. M. Eklund, A. G. James, C. McGregor, N. Percival, K. P. Smith, and D. Sow, “Real-time analysis for intensive care: development and deploymentof the artemis analytic system,” Engineering in Medicine andBiology Magazine, IEEE, vol. 29, no. 2, pp. 110–118, 2010.

[5] C. McGregor, “A cloud computing framework for real-timerural and remote service of critical care,” in Computer-Based Medical Systems (CBMS), 2011 24th International Symposium on. IEEE, 2011, pp. 1–6.

[6] G. Grimmett and D. Stirzaker, Probability and RandomProcesses, 3rded. Oxford University Press, Jul 2010.

[7] H. Khazaei, J. Mi Rightsi´c, and V. B. Right Mi Rightsi´c, “Performance analysisof cloud computing centers using M/G/m/m + r

queueing systems," IEEE Transactions on Parallel and DistributedSystems, vol. 23, no. 5, p. 1, 2012.

[8] Newling M, Fiechter RH, Sritharan L, Hoepel W, Burgsteden JA, Hak AE, et al. Dysregulated Fc gamma receptor IIa-induced cytokine production in dendritic cells of lupus nephritis patients. Clin Exp Immunol 2020; 39-199 (1) : 49.

[9] Kronbichler A, Jayne D. Response to: 'Prevention of infections in patients with antineutrophil cytoplasm antibody-associated vasculitis: potential role of hydroxychloroquine'by Novikovetal. Annals of the rheumatic diseases 2020; 79 (2):e20-e20.