Research on Implementation and Evaluation of Public Health Strategies Based on Big Data Analysis

Haoran Cheng

Xi'an Jiaotong-liverpool University, Suzhou, Jiangsu, 215000, China

1843047509@qq.com

Keywords: Big data analysis; Public health strategy; Novel coronavirus pneumonia; Application scenario

Abstract: With the global spread of the novel coronavirus pneumonia epidemic, big data analysis technology has played an important role in the formulation, implementation, and evaluation of public health strategies. However, big data analysis also faces challenges such as data quality, privacy protection, and algorithm accuracy. Based on the above background, this paper mainly studies the application scenarios and effects of big data analysis in public health strategies and the existing problems and strategies. First, this paper introduces the application examples and methods of big data analysis in epidemiological surveillance, rapid identification of suspected patients, interruption of community transmission, etc.; secondly, this paper analyzes the advantages and limitations of big data analysis in public health strategies, such as possible legal, ethical, and social issues. This paper puts forward suggestions for improving the quality and efficiency of big data analysis, including establishing unified standards, strengthening cross-departmental collaboration, and ensuring data security and privacy; finally, this paper summarizes the significance of big data analysis on public health strategies and looks forward to the future development trend.

1. Introduction

With the global spread of the novel coronavirus pneumonia epidemic, big data analysis technology has played an essential role in formulating, implementing, and evaluating public health strategies, including population monitoring, infection identification, contact tracing, etc. [1]. However, in the actual application, big data analysis also faces challenges such as data quality, privacy protection, and algorithm accuracy, so it needs to be improved.

Based on the ordinary existence of the above problems in public health, this paper mainly analyzes the application scenarios and effects of big data analysis in public health strategies based on a literature review. It discusses the existing problems and strategies based on case studies, thereby effectively improving the application level of big data analysis in the public health strategy.

This paper uses two methods: literature review and case study. Collecting and sorting out relevant literature and cases at home and abroad, it displays the application scenarios, effects, existing problems, and countermeasures of big data analysis in public health strategies, providing the research on the public health field with the necessary support.

2. Research Background of Public Health Strategy Based on Big Data Analysis

2.1 Overview of Public Health Strategies

In the research on public health strategies, the epidemiological characteristics and risk factors of diseases were mainly considered. At the same time, the socioeconomic attributes and behavioral habits of the population were ignored, which led to inaccurate formulation and evaluation of public health strategies [2]. To solve the above problems, the public health strategy is optimized by mainly combining machine learning technology and a multi-level regression model to deeply optimize the traditional single-level regression framework structure based on big data analysis to make the public health strategy more efficient and with better adaptability. Specifically, this paper uses big data analysis technology to collect and integrate information on population, environment, and medical care from multiple data sources and uses it as the independent variable and dependent variable of the multi-level regression model to explore the impact of different levels of factors on the effects of public health policies.

2.2 Application of Big Data in Public Health

In the research on public health strategies, the epidemiological characteristics and risk factors of diseases were mainly considered. In contrast, the application potential of big data analysis techniques in the field of public health was ignored, which led to the inadequacy of formulating and evaluating public health strategies [3]. Based on the above problems, a systematic study was conducted on the application of big data in public health. The research content is based on the following three key issues:

(1) How big data can improve public health monitoring and early warning capabilities;

(2) How big data can facilitate the rapid identification and detection of suspected patients;

(3) How big data can support contact tracing and community transmission interruption.

Aiming at solving these three problems, this paper conducts a comprehensive discussion from the aspects of theoretical analysis, technical methods, and practical cases to achieve a better effect of applying big data in public health strategies. Based on the analysis of the above indicators, this section conducts a detailed study on the application of big data in the field of public health. The results show that big data can improve the ability of public health staff to track and respond to infectious disease outbreaks and the ability to discover early warning signs of diseases, and the ability to research and develop diagnostic assays and therapeutics. It is of great value to research in the field of public health.

3. Research Foundation and Key Technologies of Public Health Strategy Based on Big Data Analysis

3.1 Basic Concepts and Technologies of Big Data Analysis

In the research on public health strategies, the epidemiological characteristics and risk factors of the disease were mainly considered, while the flow path and contact behavior of the crowd were ignored, which led to inaccurate public health strategies and poor prevention and control effects. Based on the above problems, an in-depth analysis of the public health strategy was conducted, mainly combining big data technology and the Markov chain model to optimize the traditional public health strategy framework structure, so that the accuracy and coverage of the prevention and control system are higher [4]. Specifically, firstly, the Hadoop platform is used to perform preprocessing operations such as cleaning, conversion, and loading of massive crowd flow logs; secondly, the Spark platform is used to perform distributed calculations on the preprocessed data to extract the flow path and contact sequence of the crowd; then, the Markov chain model is applied to predict the contact sequence of the population to obtain the transition probability matrix of the population; finally, risk assessment and intervention recommendations for each region based on the transition probability matrix are generated, and the effect is evaluated through A/B testing. The Markov Chain Model is as follows:

Set $X_1, X_2, \dots, X_n$ a sequence of random variables, if for any positive integer n and any state $i_0, i_1, \dots, i_{n+1}$

$$P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) \text{ is generated (1)}$$

Then the random variable sequence is called a Markov Chain, $P(X_{n+1} = i_{n+1} \mid X_n = i_n)$ is called the transition probability from state i_n to state i_{n+1}.

In this paper, each region or region category can be regarded as a state, and each group of people can be regarded as a Markov chain. By counting the frequency of visits by different groups of people to different regions or region categories in different time periods, the transition probability matrix can be obtained to predict which high-risk areas the population may appear in in the future or which high-risk groups the population may come into contact with.

3.2 Acquisition and Processing of Public Health Data

In the research on the novel coronavirus pneumonia, the genetic characteristics and transmission routes of the virus were mainly considered, while the mobility and contact of the population were ignored, resulting in inaccurate public health strategies and poor prevention and control effects [5]. Based on the above problems, an in-depth analysis was conducted on the novel coronavirus pneumonia. The research content was based on the key issues of crowd mobility and contact probability. Then, the risk assessment and intervention suggestion indicators were analyzed, so that the accuracy and coverage of the prevention and control system were improved. First, big data technology collects, cleans, and integrates various types of crowd flow data, including vehicles, mobile phone signals, social media, etc.; secondly, artificial intelligence technology performs pattern recognition, cluster analysis, anomaly detection, etc. on crowd flow data; and then, the Markov chain model is used to model and predict the population contact data to obtain the population transition probability matrix; finally, according to the matrix, risk assessment and intervention recommendations for each region are generated, and their effects are evaluated through A/B testing. Based on the analysis of the above indicators, this section conducts a detailed study on the novel coronavirus pneumonia The results show that big data technology and artificial intelligence technology can effectively improve the scientificity and effectiveness of public health strategies, having great value on the prevention and control of new coronavirus pneumonia.

4. Public Health Strategy Modeling and Application Based on Big Data Analysis

4.1 Construction of Public Health Policy Model Based on Big Data Analysis

This paper proposes a public health policy model construction method based on big data. This method uses big data technology to analyze the data of crowd mobility and contact, extracts feature related to the risk of disease transmission, and then uses them as the input of the machine learning model to train the model to predict the disease incidence and the effect of the intervention [6]. This method can effectively identify high-risk regions and groups and has high accuracy and real-time performance. Its specific model is as follows.

Set X as a crowd mobility and contact data set, including n samples, each sample has m features, such as means of transportation, mobile phone signal, social media, etc.; set Y as a disease incidence data set, including n samples, each sample represents the corresponding disease incidence in a region or population within a certain period of time; set Z as an intervention effect data set, which contains n samples, and each sample represents the change in the disease incidence in a corresponding region or population after implementing a certain intervention. The goal is to find a function $f(X)=Y$, which predicts disease incidence from population mobility and contact data, and a function $g(X,Y)=Z$, which predicts intervention effects from population mobility, contact, and disease incidence data.

Among them, regression model or classification model can be used to realize $f(X)$; and causal inference model can be used to realize $g(X,Y)$.

The physical meaning of each variable in the function is as follows: X Crowd mobility and contact data set: Y Disease incidence dataset: Z Intervention effect dataset: $f(X)$ Function to predict disease incidence: $g(X,Y)$ A function that predicts the effect of an intervention

4.2 Effect Evaluation of Health Policy Interventions

This section presents a method for evaluating the effects of health policy interventions based on randomized controlled trials. The method uses random allocation to intervene with the experimental group and the control group to extract the indicators related to the intervention goal, and then uses them as the input of the statistical model to train the parameters and confidence intervals for estimating the intervention effect. This method can effectively eliminate the influence of confounding factors, and identify the causal relationship in health policy interventions, having high internal validity and generalizability. Its specific model is as follows.

Set Y_i as the outcome variable observed by the i individual after the intervention; D_i is whether the i individual received the intervention (1 means yes, 0 means no); β_0 is a constant item, β_1 is the intervention effect parameter, and ϵ_i is a random error item, then:

$$Y_i = \beta_0 + \beta_1 D_i + \epsilon_i \text{ is generated (2)}$$

Where β_1 represents the average difference in the outcome variable caused by whether receiving the intervention or not when other conditions are fixed. β_1 can be estimated by the least square method or other statistical methods and its confidence interval. The physical meaning of each variable in the function is as follows: Y_i: The outcome variable observed by the i th individual after the intervention: D_i Whether the i individual received the intervention: β_0 Constant term: β_1 Intervention effect parameters: ϵ_i Random error term.

4.3 Case Study of Public Health Strategy Implementation and Evaluation

In the research on the implementation and evaluation of public health strategies, the prevention and control capabilities of infectious diseases and the supply model of medical resources were mainly considered, while the important tool of health impact assessment (HIA) was ignored, which resulted in the formulation and implementation of public health strategies are not scientific, effective, and sustainable enough. Based on the above problems, this paper conducts a case analysis of HIA. The research content is based on the theory, methods, and steps of HIA to predict, evaluate and recommend the health impact of public policies or projects in different fields and levels, so as to make the implementation and evaluation of public health strategies more comprehensive, reasonable and beneficial to human health. Paragraphs use the first, second, then, and last context word order in the turning process. Finally comes the concluding paragraph.

Based on the analysis of the above indicators, this paper conducts a detailed study on HIA. The results show that HIA is an effective public health decision support tool that can help identify and reduce adverse health factors, promote multi-sectoral cooperation and social participation, and improve the quality and effectiveness of strategy implementation and evaluation of public health. It is of great value to the construction and development of China's public health system.

4.4 Prospects for the Implementation and Evaluation of Public Health Strategies Based on Big Data Analysis

This section proposes a public health strategy implementation and evaluation method based on big data analysis. This method uses social networks and search engines to mine public health behaviors and needs, extracts indicators related to the influencing factors and effect evaluation of public health policies or projects, and then uses them as the input of machine learning models to train the classifier to identify the health risks and needs of different groups of people, and proposes corresponding policy recommendations or interventions based on the prediction results. This method can effectively analyze the health problems and inequalities existing in society and has high accuracy and real-time performance. The specific model is as follows: X represents the user data obtained from social networks and search engines, including the user's basic information, health status, behavior habits, topics of concern, etc.; Y represents the category of the user to which the user belongs, such as age group, gender, region, income level, etc.; Z represents the health risks or needs faced by users, such as chronic diseases, infectious diseases, psychological stress, health care knowledge, etc.; f represents machine learning models, such as support vector machines, random forests, neural networks, etc.; g represents policy recommendations or interventions generated based on Z, such as publicity and education, medical services, material subsidies, etc.

5. Conclusion

In the related research on epidemiological monitoring, the transmission model of infectious diseases was mainly considered, while the movement pattern of the crowd was ignored, which led to the problem of inaccurate epidemic prediction. Based on the above problems, the epidemiological monitoring method was improved for the novel coronavirus pneumonia epidemic by combining big data analysis technology and machine learning model to deeply optimize the traditional SIR framework structure, so that the prediction effect of the number of infected people and mortality of novel coronavirus pneumonia is improved.

References

[1] J. Yang, Y. Li, Q. Liu, L. Li, A. Feng, T. Wang, S. Zheng, A. Xu, and J. Lyu, "Brief introduction of medical database and data mining technology in big data era," Journal of Evidence-Based Medicine, vol. 13, pp. 57-69, 2020.

[2] J. W. Lee, "Big data strategies for government, society and policy-making," Lee, Jung Wan (2020). Big Data Strategies for Government, Society and Policy-Making. Journal of Asian Finance Economics and Business, vol. 7, pp. 475-487, 2020.

[3] N. L. Bragazzi, H. Dai, G. Damiani, M. Behzadifar, M. Martini, and J. Wu, "How big data and artificial intelligence can help better manage the COVID-19 pandemic," International journal of environmental research and public health, vol. 17, p. 3176, 2020.

[4] S. Dash, S. K. Shakyawar, M. Sharma, and S. Kaushik, "Big data in healthcare: management, analysis and future prospects," Journal of Big Data, vol. 6, pp. 1-25, 2019.

[5] C. J. Wang, C. Y. Ng and R. H. Brook, "Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing," Jama, vol. 323, pp. 1341-1342, 2020.

[6] Y. Wang, L. Kung and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," Technological forecasting and social change, vol. 126, pp. 3-13, 2018.