ISSN:2790-1688 Volume-14-(2025)

Regression-based Bibliometric Studies Based on Citespace

Jia Peng*

Applied Statistics, School of Statistics, Renmin University of China, Beijing, China

pengjia1210@ruc.edu.cn

Abstract. This paper employs CiteSpace software to conduct bibliometric and visualization analyses on the literature related to "regression" from the Web of Science Core Collection, spanning the period from 1975 to 2024, in order to reveal the current research status, hotspots, and trends in the field of regression analysis. The findings indicate that the number of publications in this field has shown a stable growth trend, with close international cooperation, and the United States, China, the United Kingdom, and other countries occupying important positions in this field. Keyword analysis shows that research hotspots include multicollinearity, machine learning, nonparametric regression, quantile regression, etc., and in recent years, the integration of machine learning and regression analysis has become increasingly close and has become a research frontier. This paper provides a comprehensive analysis of the current research status and future research directions for researchers in the field of regression analysis..

Keywords: Regression analysis; CiteSpace; Bibliometrics; Research hotspots; Research trends.

1. Introduction

As one of the most basic and important methods in statistics, regression analysis can be traced back to the 19th century. Francis Galton put forward the concept of "regression" for the first time when studying the transmission of characteristics in genetics, which was used to describe the phenomenon of "regression" of offspring characteristics to the average value [1]. Subsequently, Karl Pearson et al. [2] systematically developed regression theory and laid the foundation of modern regression analysis. Since the 20th century, with the progress of computer technology and the continuous improvement of statistical theory, regression analysis has gradually expanded from simple linear model to multiple regression, nonlinear regression, generalized linear model and more complex machine learning methods. Regression analysis aims to reveal the law behind the data by establishing the relationship model between the dependent variable and one or more independent variables and to predict and explain the phenomenon [3]. Its application covers almost all disciplines: in economics, regression is used to analyze market trends and predict economic indicators [4-5]. In medicine, it is used to study the relationship between diseases and risk factors [6-7]. In social science, regression analysis helps researchers understand the driving factors of human behaviour and social phenomena [8-9]. In addition, regression analysis also provides theoretical support for the development of machine learning and artificial intelligence [10]. Linear regression is the cornerstone of neural networks and support vector machines, while logistic regression plays a key role in classification.

CiteSpace is a bibliometric analysis software developed by Professor Chen Chaomei, which is specially designed to study the knowledge structure and dynamic evolution of academic literature [11]. Analyzing a large number of academic literature data generates a visual knowledge map, which helps researchers quickly identify research hotspots, track the development of disciplines, and reveal structural changes in knowledge fields. The core functions of CiteSpace include author cooperation network analysis, co-citation analysis and keyword co-occurrence analysis, which can intuitively show the relevance of academic cooperation mode, high-impact literature and research topics. Its powerful data processing ability and visualization effect provide researchers with a comprehensive analytical perspective. Searching for articles with CiteSpace in the title on the Web of Science (WoS) website, we can find that CiteSpace is in General Internal Medicine [12-13] (252 articles), Information Science Library Science [14-15] (248 articles), Environmental Sciences Ecology [16-17] (247 articles) and other disciplines have been widely used. However, although CiteSpace has

demonstrated its powerful analytical ability in many disciplines, its application in the field of statistics is relatively blank, especially in the study of regression analysis, a classical statistical method.

The aim of this study is to analyze the research status, hot spots and trends in the field of regression by using CiteSpace software and the WoS database as data sources. By using "regression" as the keyword, the period is set to 1975-2024, the Article type is limited to "article" and the language is "English", a total of 1,331,896 documents are retrieved, and then the quota is set according to the proportion of published articles per year, and 5000 effective papers are selected according to the relevance. In this paper, a variety of bibliometrics and visual analysis methods will be adopted, and the selected literature will be analyzed by CiteSpace5.7.R5 software, to explore the research survey, core research topics and frontier trend statistics in this field, and draw the annual publication volume, authors, institutions, countries or regions cooperation knowledge map to understand the research survey in the field of regression. By analyzing the knowledge map of keyword co-occurrence and clustering, this paper reveals the research hotspots, and explores the trend frontier of this research field with the help of keyword timeline and keyword emergence, to systematically comment on the research in this field and provide useful reference for subsequent related research.

2. Literature Review

As the core tool of statistical modelling, the theory and method of regression analysis have been constantly developed to meet the increasingly complex data analysis needs, and the papers reviewing regression analysis have also been constantly updated and developed. In 1990, in the analysis of biochemical data, Leatherbarrow [18] discussed the advantages and disadvantages of linear and nonlinear regression, revealed the limitations of linear regression on the assumption of error distribution, and proposed the advantages of nonlinear regression in dealing with complex equation fitting, but its computational complexity and sensitivity of initial parameters limited its application scope. In 1998, Yu and Jones [19] reviewed the quantile regression and proposed a local linear dual-core smoothing method. By optimizing the bandwidth selection, this method significantly improves the performance of quantile regression in dealing with heteroscedasticity and complex curves and provides a more robust solution. With the increased data complexity, traditional regression methods face challenges in dealing with fuzziness and uncertainty. In 2001, Chang and Ayyub [20] systematically compared the differences between fuzzy regression and traditional least square regression, reviewed three methods of minimizing fuzziness, least square method and interval regression analysis, and further put forward mixed fuzzy least square regression, which provides a new idea for integrating randomness and fuzziness.

In 2005, aiming at sparse longitudinal data, Yao et al.[21] proposed a nonparametric function linear regression method. Through functional principal component analysis and conditional expectation estimation, the prediction accuracy is significantly improved and the application range of the functional determination coefficient was expanded. In solving the problem of multicollinearity and over-fitting, in 2009, McDonald [22] reviewed the ridge regression method and revealed the relationship between the ridge trace map and data underlying structure by analyzing the influence of ridge parameter change on coefficient, which provides theoretical support for model optimization. In 2014, Loh [23][22] reviewed the development of classification and regression trees in recent 50 years, and pointed out that regression trees can fit many traditional statistical models, such as the least square method, quantile regression, logistic regression and so on, as well as longitudinal and multi-response data models, and combed the core ideas of its main algorithms. In 2018, Ranstam and Cook[24] reviewed the LASSO regression, which realized variable selection and model simplification through L1 norm constraint, and provided an efficient tool for high-dimensional data analysis. In recent years, the regression method has been used in machine learning, and the research on optimization and expansion of this method continues to deepen. In 2020, Maulud and Abdulazeez [25] summarized the research progress of simple regression, multiple regression and polynomial regression in machine learning, and emphasized the correlation between model efficiency evaluation and data set

characteristics. In 2025, Kock and Klein [26] proposed a multivariate distribution regression model based on Gaussian copula, which provided a new idea for joint modelling of high-dimensional response through Bayesian inference and contraction prior. Compared with the traditional literature review, the visual analysis of CiteSpace enables us to understand the development and future trend of regression methods more clearly, which lays a solid foundation for further research in this field.

3. The Statistics of the Annual Publication Output

The annual publication output can reflect the overall situation of a research field in a certain period, and objectively show the research process and development law of this research field [27]. Figure 1 shows the annual publication output of literature related to regression retrieved from the WoS database. It can be seen that the research on the regression field in the international scope shows a steady growth trend, and the growth rate has accelerated since 2000. In the past two years, the number of papers published has decreased compared with that in 2022, but it remains at a high level.

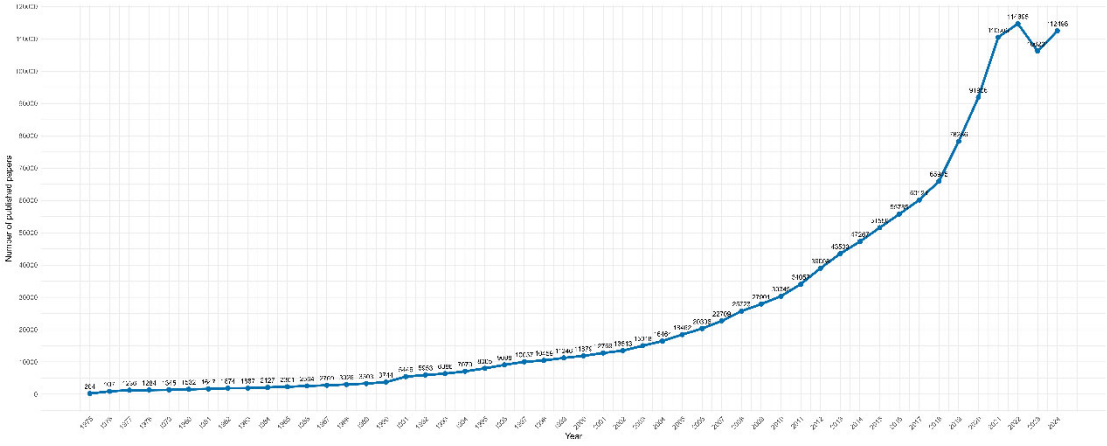

Figure 1 The Statistics of the Annual Publication Output on WOS

4. The Publication Output Analysis of the Countries, Regions and Institutions

Under the current background of globalization, international cooperation plays a vital role in scientific research and technological innovation. The top five countries/regions in the number of published papers are the United States, China, Britain, Germany and Canada, and the intermediary centrality of the top three countries is higher than 0.4, indicating that they occupy an important position in this field. In addition, as shown in Figure 2, many cooperation networks have been formed around the United States, China, Britain and other countries, and have spread all over the world, indicating that good international cooperation relations have been formed in this field. Furthermore, by analyzing the cooperation map of the publishing institutions (Figure 3), we can understand the status and role of various research institutions in the global scientific research network in more detail. It can be seen that the Chinese Academy of Sciences (71) and the University of Minnesota (67) rank in the top two, and the relevant research started earlier. The University of São Paulo in Brazil, the University of Leuven in Belgium and Korea University also occupy an important position in this Figure. The research institutions cooperate closely and have established relatively close cooperation, which is helpful for cross-regional scientific research exchange and cooperative innovation in the field of regression.

5. The Author and the Cited Author Analysis

By analyzing the authors, we can understand the representative authors and core groups in this field and explore the cooperative relationship between the authors. As can be seen from Figure 4(a),

many scholars study the regression field, and they have formed a dense cooperation network led by scholars such as HENG LIAN, GAUSS M CORDEIRO, EDWIN M M ORTEGA, which shows that these scholars have conducted frequent cooperation and exchanges, formed many research teams, and promoted the in-depth development of related fields. In addition, some scholars have built small cooperation networks, which shows that there are also close exchanges and cooperation between them. However, some cooperative groups are small, consisting of only two authors, and even a few scholars choose to carry out research alone without forming a cooperative team. This phenomenon limits the possibility of cross-disciplinary in-depth research to some extent. In Figure 4(b), we found some highly cited authors and their research results had a far-reaching impact on this field. For example, the literature of COX DR, FRIEDMAN JH and KOENKER R is widely cited, and it contains classic theories and important achievements in this field. The research contents of these highly cited authors cover statistics, machine learning and other aspects, which provide a solid theoretical basis and technical support for the follow-up research. Meanwhile, we can also see some emerging cited authors, such as R CORE TEAM, YUAN M, ZOU H, etc., and their research results have gradually attracted attention and shown strong influence in some sub-fields.

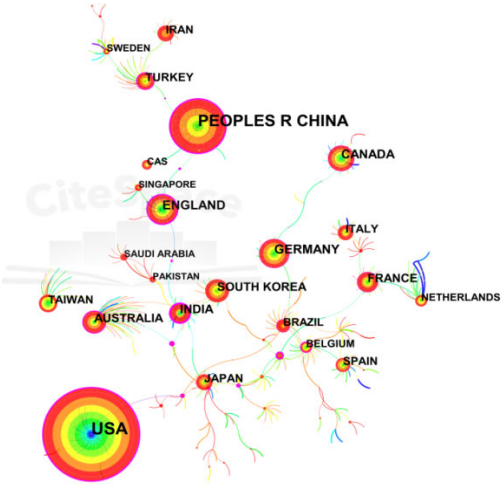

Figure 2 Map of Cooperation between Countries or Regions

Figure 3 Map of Institutional Cooperation

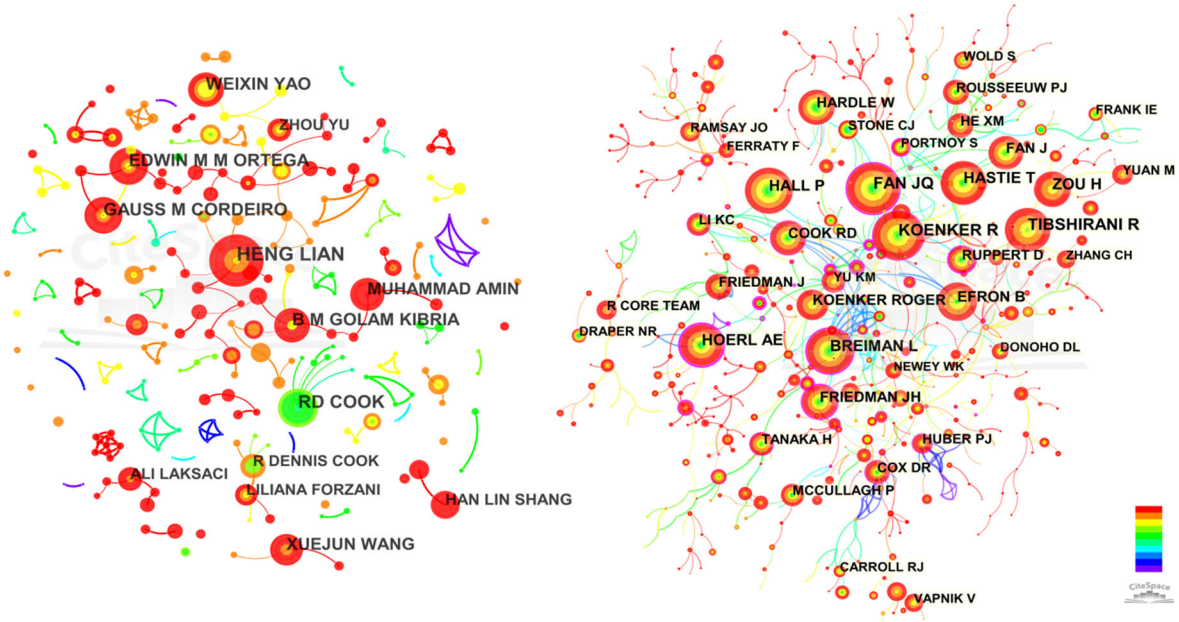

Figure 4 Cooperative Atlas of Author (a) and Cited Author (b)

6. Keyword Analysis

6.1 Keyword co-occurrence knowledge map

Keywords are highly concise and summarise the content of the article, and the knowledge map of keyword co-occurrence can reveal the research hotspots and topic associations by showing the frequency and relationship of different keywords appearing together in the same paper [28]. After importing the literature into Citespace software, we select “Keyword” as the node type, set the threshold to Top50, and set the time slice to 5 years. After running, a keyword co-occurrence map consisting of 157 nodes and 427 connections is obtained (Figure 5). Wherein each node represents a keyword; node size represents the frequency of keyword occurrence; the colour of the node represents different times, and the thicker the colour circle in the node, the higher the frequency of the keyword appearing at the corresponding time of the colour. The number of connecting lines indicates the co-occurrence coefficient of keywords, and the more connecting lines represent the closer correlation between keywords. The top ten keywords in frequency and agency centrality are shown in Table 1. High frequency not only represents the heat of research but also reflects continuous attention and in-depth research. The centrality of network nodes shows the bridge function of keywords in the network [29], “multiple regression” ranks first with 0.49, which shows that it plays an important role in connecting different research topics.

6.2 Keyword Clustering Diagram

Citespace can identify the internal association of each group of keywords, automatically generate the most representative category labels accordingly, and classify keywords with similar themes or contents into groups. It can not only realize the systematic classification of research data but also effectively identify the focus issues and development trends in the subject field. In CiteSpace clustering results, the average contour value (S) and the clustering module value (Q) are two important evaluation indexes. S represents the compactness of samples in the cluster. It is generally believed that when S is greater than 0.7, the clustering result is convincing. When Q is greater than 0.3, the divided community structure is significant [30]. In this paper, the LLR algorithm is used to extract clustering labels with keywords as the identification, and the minimum clustering variable is set to 10 to form a keyword clustering map (Figure 6), and the detailed results of 8 clusters are listed (Table 3). The clustering result S is 0.9195, greater than 0.7, and Q is 0.7564, greater than 0.3, which

indicates that the clustering result is convincing and the network community structure is remarkable. The keyword clustering results are analyzed as follows:

Cluster #0: The topic is multicollinearity. This clustering mainly focuses on multicollinearity problems and their solutions. Keywords include “ridge regression”, “principal component regression”, “regression model”, “ridge estimator”, “quantile regression” and so on. These keywords show that researchers have made a lot of explorations in dealing with the high correlation of variables, especially through ridge regression and principal component regression to alleviate the multicollinearity problem.

Cluster #1: The topic is regression. Clustering covers a variety of regression models and related technologies. Keywords include “machine learning”, “robust regression”, “outliers”, “consistency”, “random forest” and so on. This shows that in addition to the traditional regression methods, modern technologies such as machine learning and random forest have also been widely used in regression analysis. In addition, “robust regression” and “outliers” reflect researchers’ concern about outliers and robust regression.

Cluster #2: The topic is optimization. This clustering mainly involves various optimization methods and technologies, including “sliced inverse regression”, “quantile regression”, “simulation”, “gaussian process regression” and “neural network”. These keywords show researchers’ efforts in optimization algorithms, simulation technology and neural networks.

Cluster #3: The topic is regression analysis. This cluster covers many aspects of regression analysis, including “fuzzy regression”, “least squares”, “quantitative regression”, “local linear” and so on. This shows that researchers not only pay attention to the continuous upgrading of traditional regression methods but also explore new methods such as fuzzy regression and local linear regression to meet different analysis needs.

Cluster #4: The topic is nonparametric regression. The clustering focuses on nonparametric regression methods and related technologies, including “linear regression”, “asymptotic normality”, “nonlinear regression”, “regression spline”, “goodness-of-fit” and so on. These keywords indicate that researchers have conducted extensive research in the field of nonparametric regression.

Cluster #5: The topic is support vector regression. Its related sub-clusters include “support vector machine”, “twin support vector regression”, “multicollinearity”, “linear programming”, “kernel methods” and so on. This shows that support vector machine regression plays an important role in dealing with multicollinearity and linear programming problems.

Cluster #6: The topic is quantile regression. This cluster discusses the related topics of quantile regression, including “poisson regression”, “logistic regression”, “median regression”, “multicollinearity”, “ridge regression” and so on. This shows that quantile regression is not only applied in median regression, but also expanded in other regression types such as poisson regression and logistic regression, and is closely related to multicollinearity and ridge regression.

Cluster #7: The topic is dimension reduction. This clustering mainly involves the dimensionality reduction method and its application. Keywords include “central subspace”, “sliced inverse regression” and so on. This shows that researchers have conducted in-depth research in the field of dimensionality reduction, especially through central subspace and slice inverse regression to deal with the analysis of high-dimensional data and functional data.

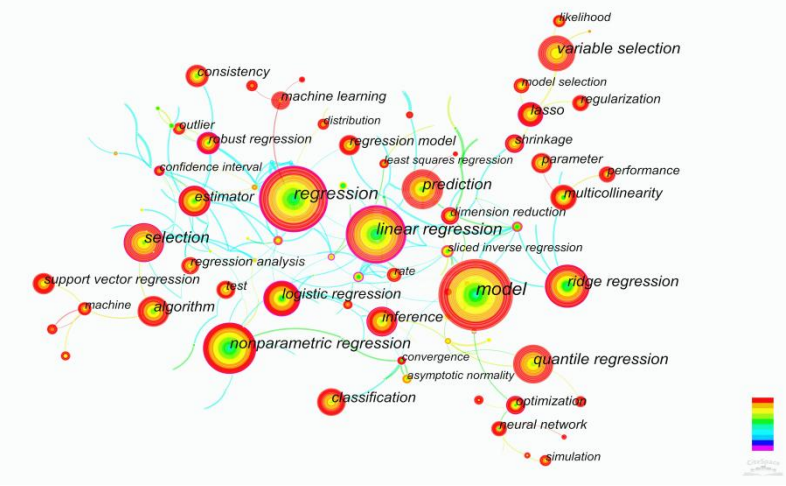

Figure 5 Keyword Co-occurrence Network Map

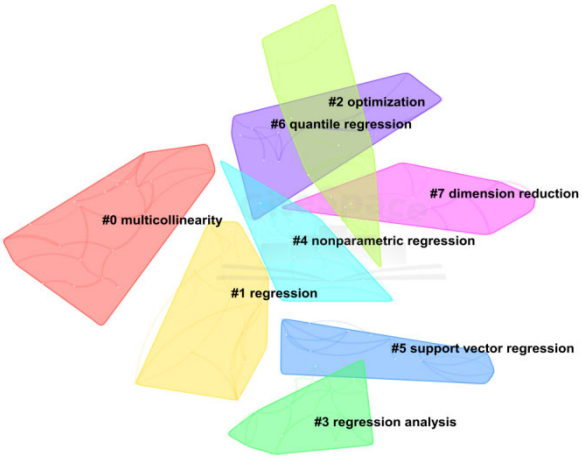

Figure 6 Keyword Clustering Map

Table 1 Top 10 Keywords by Frequency of Occurrence and Top 10 Keywords by Intermediary Centrality

Ranking	Keyword	Frequency	Keyword	Intermediary centrality
1	model	717	multiple regression	0.49
2	regression	561	kernel regression	0.43
3	linear regression	427	regression	0.38
4	selection	301	partial least square	0.38
5	nonparametric regression	287	generalized linear model	0.37
6	quantile regression	271	sliced inverse regression	0.33
7	variable selection	270	projection pursuit regression	0.27
8	prediction	262	logistic regression	0.26
9	ridge regression	240	linear regression	0.25
10	logistic regression	200	variable subset selection	0.24

Table 2 Keyword Clustering Information

Serial number	Cluster name	Clustering sub-cluster
#0	multiconllinearity	ridge regression; principal component regression; regression model; ridge estimator; Quantile regression, etc

#1	regression	machine learning; robust regression; outliers; consistency; random forest et al.
#2	optimization	sliced inverse regression; quantile regression; simulation; gaussian process regression; Neural network, etc
#3	regression analysis	fuzzy regression; least squares; quantile regression; local linear; Regression et al.
#4	nonparametric regression	linear regression; asymptotic normality; nonlinear regression; regression spline; Goodness-of-fit, etc
#5	support vector regression	support vector machine; twin support vector regression; multicollinearity; linear programming; Kernel methods, etc
#6	quantile regression	poisson regression; logistic regression; median regression; multicollinearity; Ridge regression et al.
#7	dimension reduction	central subspace; sliced inverse regression; regression graphics; model; Functional data analysis, etc

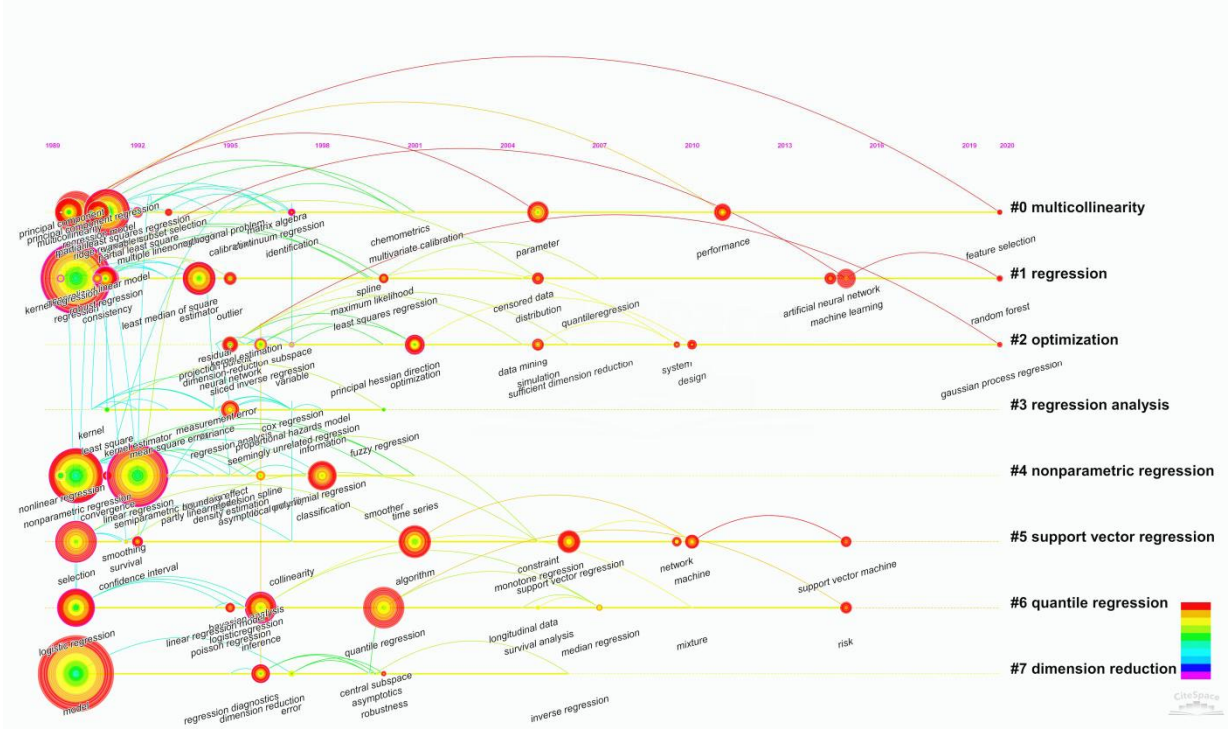

Figure 7 Keyword Timeline Atlas

6.3 Keyword timeline chart

Through the keyword timeline chart, we can clearly see the popularity and evolution of keywords at different time points, reveal the development trend, hot spot transfer and theme rise and fall in the research field, and look forward to the possible research direction in the future. As can be seen from Figure 7, the time evolution of keywords presents prominent stage characteristics. From the end of the 1980s to the beginning of the 1990s, the research mainly focused on basic regression methods such as “principal component regression” and “kernel regression”. As time goes on, keywords gradually shift to more complex and advanced methods. In the 2000s, keywords such as “support vector machine” and “neural network” began to appear and heated up rapidly, reflecting the wide application of machine learning and artificial intelligence technology in regression analysis. During this period, nonparametric, quantitative and other nonparametric and quantile regression methods have gradually attracted attention, showing the researchers’ understanding of the limitations of traditional regression models and their exploration of new methods. After 2010, keywords such as “multicollinearity” and “dimension reduction” have become new research hotspots, which shows that

multicollinearity and dimension reduction are increasingly important in high-dimensional data processing. In addition, the popularity of machine learning algorithms such as “machine learning” and “random forest” continues to rise, indicating that future research will pay more attention to applying data-driven and intelligent methods. It is worth noting that some classic keywords, such as “least squares regression” and “logistic regression”, still maintain a certain degree of attention, which shows that these basic methods still have irreplaceable value in practical application.

6.4 Keyword pop-up diagram

Through keyword pop-up analysis, we can identify and explore the research frontiers and latest trends in a certain field. On the basis of the frequency of keyword emergence, hot words are determined according to the growth rate of keyword occurrence times. The correlation characteristics between these hot words and time are usually regarded as the research frontier in a certain field [31]. Figure 8 is a map of the top 25 emergent words in the literature on regression from 1975 to 2024. As can be seen from Figure 8, nonparametric regression has shown a strong citation explosive force since 1990, and continued until 2014, with its citation intensity as high as 15.72. The research in this period mainly focuses on dealing with the nonlinear relationship and complex structure in data, and the flexibility and adaptability of the model are improved by regression spline and kernel method. Nonlinear regression is similar to nonparametric regression, showing a significant citation burst in the same period, with a citation intensity of 14.63. Researchers use various nonlinear models and techniques to solve the problems that traditional linear regression cannot handle, especially in the fields of biomedicine, engineering and social science. Secondly, least square, multiple regression, partial least square and principal component regression also experienced a citation outbreak from 1990 to 2014. These methods are mainly used to deal with the linear relationship, the influence of multiple independent variables and the multicollinearity problem in high-dimensional data. The accuracy of prediction and the stability of the model are improved by a multivariate linear model and dimension reduction technology. After entering 1995, error, sliced inverse regression, polynomial regression, calibration and fuzzy regression began to show strong citation explosiveness. These methods have played an important role in dealing with complex data structures, nonlinear relationships, model calibration and uncertainty, reflecting researchers’ concern about model accuracy and robustness. After entering the 21st century, central subspace, kernel regression, asymptotic normality, maximum likelihood, median regression, survival analysis and survival showed a significant citation explosion from 2000 to 2019. These methods have played an important role in processing high-dimensional data, nonparametric estimation, large sample properties, robust statistics and survival data analysis, and further promoted the development and application of regression analysis methods. Finally, classification, machine learning, gaussian process regression, random forest and feature selection showed strong citation explosiveness from 2015 to 2024. Among them, machine learning has become the most concerned research hotspot in recent years with the highest citation intensity of 31.14, which reflects the deep integration of statistics and machine learning in the era of big data. To sum up, the keyword pop-up diagram not only reveals the research hotspots and development trends in the field of regression analysis but also provides an important reference for future research. With the continuous progress of big data and artificial intelligence technology, regression analysis methods will continue to innovate and develop, providing more powerful tools and support for solving complex real problems. From the traditional linear regression to the modern machine learning method, the field of regression analysis has shown vigorous vitality and broad application prospects.

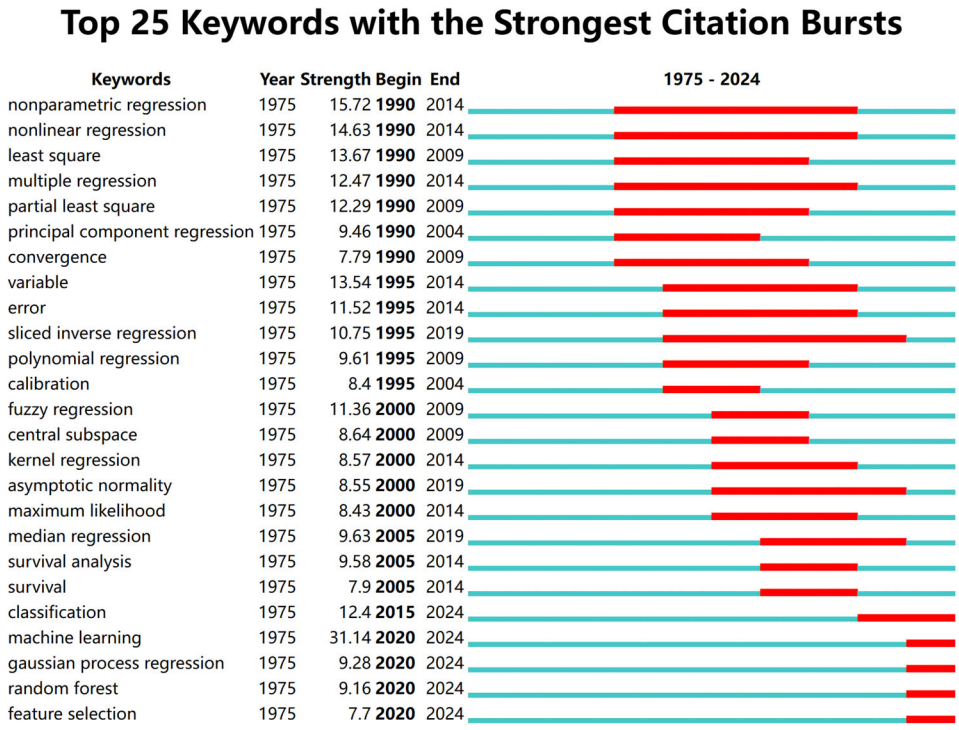

Figure 8 The Keyword Emergence Map

References

[1] Galton, F. (1886). Regression towards mediocrity in hereditary stature. The Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263.

[2] Pearson, K. (1905). On the general theory of skew correlation and non-linear regression (No. 14). Dulau and Company.

[3] Wang, G. C., & Jain, C. L. (2003). Regression analysis: modelling & forecasting. Institute of Business Forec.

[4] Farimani, S. A., Jahan, M. V., Fard, A. M., & Tabbakh, S. R. K. (2022). Investigating the informativeness of technical indicators and news sentiment in financial market price prediction. Knowledge-Based Systems, 247, 108742.

[5] Kumbure, M. M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. Expert Systems with Applications, 197, 116659.

[6] Yang, X., Li, K., Wen, J., Yang, C., Li, Y., Xu, G., & Ma, Y. (2024). Association of the triglyceride glucose-body mass index with the extent of coronary artery disease in patients with acute coronary syndromes. Cardiovascular diabetology, 23(1), 24.

[7] Li, W., Shen, C., Kong, W., Zhou, X., Fan, H., Zhang, Y., ... & Zheng, L. (2024). Association between the triglyceride glucose-body mass index and future cardiovascular disease risk in a population with Cardiovascular-Kidney-Metabo lic syndrome stage 0–3: a nationwide prospective cohort study. Cardiovascular Diabetology, 23(1), 292.

[8] Piza, E. L., Connealy, N. T., Sytsma, V. A., & Chillar, V. F. (2023). Situational factors and police use of force across micro-time intervals: A video systematic social observation and panel regression analysis. Criminology, 61(1), 74-102.

[9] Cava, L. L., Aiello, L. M., & Tagarelli, A. (2023). Drivers of social influence in the Twitter migration to Mastodon. Scientific Reports, 13(1), 21626.

[10] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(2), 140-147.

[11] Chen, C. (2016). CiteSpace: a practical guide for mapping scientific literature (pp. 41-44). Hauppauge, NY, USA: Nova Science Publishers.

[12] Zhou, Q., Kong, H. B., He, B. M., & Zhou, S. Y. (2021). Bibliometric analysis of bronchopulmonary dysplasia in extremely premature infants in the Web of Science database using CiteSpace software. Frontiers in Pediatrics, 9, 705033.

[13] Liu, Y., Dong, Y., Zhou, W., & Yu, J. (2025). Visual analysis of emerging topics and trends in contrast agent extravasation research in medical imaging: a bibliometric study using CiteSpace and VOSviewer. Frontiers in Medicine, 12, 1472637.

[14] Li, P., Yang, G., & Wang, C. (2019). Visual topical analysis of library and information science. Scientometrics, 121(3), 1753-1791.

[15] Song, Y., Wei, K., Yang, S., Shu, F., & Qiu, J. (2023). Analysis of the research progress of library and information science since the new century. Library hi tech, 41(4), 1145-1157.

[16] Zhang, Y., Li, C., Ji, X., Yun, C., Wang, M., & Luo, X. (2020). The knowledge domain and emerging trends in phytoremediation: a scientometric analysis with CiteSpace. Environmental Science and Pollution Research, 27, 15515-15536.

[17] Wu, L., Miao, H., & Liu, T. (2024). Development in agricultural ecosystems' carbon emissions research: A visual analysis using CiteSpace. Agronomy, 14(6), 1288.

[18] Leatherbarrow, R. J. (1990). Using linear and non-linear regression to fit biochemical data. Trends in Biochemical Sciences, 15(12), 455-458.

[19] Yu, K., & Jones, M. (1998). Local linear quantile regression. Journal of the American Statistical Association, 93(441), 228-237.

[20] Chang, Y. H. O., & Ayyub, B. M. (2001). Fuzzy regression methods–a comparative assessment. Fuzzy Sets and Systems, 119(2), 187-203.

[21] Yao, F., Müller, H. G., & Wang, J. L. (2005). Functional linear regression analysis for longitudinal data.

[22] McDonald, G. C. (2009). Ridge regression. Wiley Interdisciplinary Reviews: Computational Statistics, 1(1), 93-100.

[23] Loh, W. Y. (2014). Fifty years of classification and regression trees. International Statistical Review, 82(3), 329-348.

[24] Ranstam, J., & Cook, J. A. (2018). LASSO regression. Journal of British Surgery, 105(10), 1348-1348.

[25] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. Journal of applied science and technology trends, 1(2), 140-147.

[26] Kock, L., & Klein, N. (2025). Truly multivariate structured additive distributional regression. Journal of Computational and Graphical Statistics, 1-13.

[27] Wang, W., & Lu, C. (2020). Visualization analysis of big data research based on Citespace. Soft Computing, 24(11), 8173-8186.

[28] Zhong, D., Li, Y., Huang, Y., Hong, X., Li, J., & Jin, R. (2022). Molecular mechanisms of exercise on cancer: a bibliometrics study and visualization analysis via CiteSpace. Frontiers in Molecular Biosciences, 8, 797902.

[29] Yuan, X., & Lai, Y. (2023). Bibliometric and visualized analysis of elite controllers based on CiteSpace: landscapes, hotspots, and frontiers. Frontiers in Cellular and Infection Microbiology, 13, 1147265.

[30] Chen, Y., Chen, C., Liu, Z., Hu, Z., & Wang, X. (2015). Methodological functions of CiteSpace knowledge graphs. Stud. Sci. Sci, 33(02), 242-253.

[31] Geng, Y., Zhang, N., & Zhu, R. (2023). Research progress analysis of sustainable smart grid based on CiteSpace. Energy Strategy Reviews, 48, 101111.