Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 1, No. 2, 2022

Analysis of National Social Science Foundation Projects Based on LDA Thematic Analysis

Mingjia Wu, Yanqi Wang, Wenhui Ding

School of Management Science and Engineering, Anhui University of Finance and Economics, Bengbu, 233030, China

Abstract: The National Social Science Foundation (NSSF) projects are the only national-level philosophical and humanities social science research projects in China at present, which have an important role in promoting the progress of China's economy and society and the development of humanities and social sciences. In order to understand the direction of the evolution of the main body of social science foundation projects and provide reference for related scholars, the LDA main body analysis method was applied to analyze the social science foundation projects from 2011 to 2020 as a research sample, and the results show that the research as a whole is focused on the keywords of "China, mechanism, path, era, village, culture and community". "The three major high frequency modules are the new development pattern of domestic and international "double cycle", rural revitalization and digital governance research.

Keywords: LDA thematic analysis, Social Science Foundation.

1. Research Background

1.1. Development

With the advent of artificial intelligence and the era of big data, the network and information technology have gradually penetrated into every aspect of life and work, real-world textual information is more often presented electronically, and text mining has become a research hotspot and learning focus in the information field. The use of computers to realise the recognition and analysis of massive amounts of text has become a popular topic of research.

With the continuous development of the digital economy, the value of data has gradually emerged and become an important driving force for social development. The National Social Science Foundation of China (NSFC) is the highest level of research funding in the humanities and social sciences, with strong authority and representativeness. By carrying out text mining on the selected topics of the National Social Science Foundation projects, we can have high-value topics in terms of academic value, application value, novelty and project data, etc., and finally determine the project topics, which will greatly reduce the problem of "difficult topic selection".

1.2. Current Situation and Development Trend of Domestic and Foreign Research

In the late 1950s, H.P. Luhn conducted pioneering research in this field and proposed the idea of word frequency statistics for automatic classification. 1960, Maron published In 1960, Maron published the first paper on automatic classification, followed by K. Spark, G. Salton and K.S. Jones, who also carried out fruitful research in this area. At present, text mining research abroad has moved from the experimental stage to the practical stage, with project evaluation agencies in the US mainly including Congress, unofficial agencies and the universities themselves. Expert review and econometrics are the main methods used in the evaluation of research in American universities. The main methods are based on quantitative indicators, studying the corresponding data of various research results and analysing the impact of quantitative data on research. With scientific research activities showing multi-dimensional penetration of disciplines and the increasingly integrated structure of the connotation and extension of knowledge systems, the establishment and improvement of the evaluation system of colleges and universities has been the first to bear the brunt.

The evaluation of research projects in China started late, and with the continuous development of science and technology, scientific and technological innovation has gradually become the main driving force of economic and social development. Scientific research projects have been highly valued by all sectors of society and the funds invested in them have been growing. The reform of the scientific research system and mechanism, especially from the 1990s to the present, has gradually brought attention to the evaluation of scientific research projects. With the advent of the fourth industrial revolution, China has put forward the concept of "Made in China 2025", which has placed even higher demands on the country's scientific research. Research evaluation plays a key role in scientific research activities and the training of human resources. The evaluation methods used in China are similar to those used abroad, with qualitative analysis based on peer review and quantitative analysis based on metrological methods, and there is still a gap between them in terms of index systems, standardisation and scientificity, which is related to the fact that the development of scientific research evaluation in China is still at a perfect stage.

In recent years, there has been a proliferation of research on TDT technology at home and abroad, with language models dominating in text-based topic discovery, where LDA uses generative model thinking to achieve topic assignment and has achieved many results in application. Li Chang introduced the technical word context IPC to form the WI-LDA model, which achieved better results in topic generation for the assigned task. Tan Xu fused the ARMA model and LDA for dynamic presentation and fine-grained segmentation in sentiment analysis. In addition, a topic count K needs to be determined as a control variable in order to be able to compare the application of the experiment more intuitively. Blei used the Perplexity metric to measure the uncertainty of the model's assignment of topic models, but this approach has a

tendency to favour high topic counts. Wang and others use the smallest curve inflection point as the number of topics based on Occam's razor criterion, but this method lacks stability and is difficult to ensure that the resulting solution is optimal; Griffiths uses the standard Bayesian statistical method with a log marginal likelihood function instead of the perplexity metric; Guan and Peng use JS scatter to calculate the variance of each topic-word distribution parameter around its The combination of JS scatter to calculate the magnitude of the variance of each topic-word distribution parameter around its mean and the perplexity metric proposed the Perplexity-Var metric. Teh proposed the Hierarchical Dirichlet Process Hdp, in which each sample is independently drawn from the mixture distribution and the final mixture score is generated by completing the sampling process, and the experimental results found that the best mixture score is consistent with the results obtained by the perplexity method.

2. Research Methods

2.1. Python Data Crawling Mining Method

Firstly, based on python, we crawled and sorted out the relevant literature of China's social science research fund projects, and used the jieba library to sort, organize, sort and cluster the short texts of the research topics to form the data base for this project.

2.2. LDA theme extraction method

LDA is a model for inferring the topic distribution of documents, proposed by Blei et al. in 2003.

The topic model is a multi-layer Bayesian generative model with the term "document-a-topic". At its core, it considers a document.The LDA Topic Model is a multi-layered Bayesian generative model of the 'one topic per document, one word' model, which considers that a document contains multiple topics and that each word is generated by a fixed topic. The LDA topic model can transform textual information into numerical information, so that the topic of each document in the document set is given in the form of a probability distribution, and by extracting the topics in the document, it is possible to achieve topic clustering or text classification according to the topic distribution.

3. Research Results

3.1. The Overall Status of the National Social Science Foundation (Nssf) Projects

The National Social Science Foundation is divided into 23 disciplines, and the discipline settings remain basically stable in each year. According to the list of projects for each year downloaded from the website of the National Office of Philosophy and Social Science Planning, the number of projects established in each discipline in each year was statistically summarized, and the hot situation and development of the NSF projects could be obtained.

The total number of projects of the National Social Science Foundation has shown a steady growth over the past 10 years, from 2,883 in 2011 to 4,625 in 2020, an increase of 60.42%. In terms of individual disciplines, the number of projects in some disciplines fluctuates slightly in individual years, but in terms of the overall trend over the 10-year period, all disciplines have shown growth. The increase in the number of projects in each discipline varied considerably, ranging from 15.48% to 256.52%. The disciplines with the highest increases included Archaeology, Marxism-Leninism-Science and Society, Statistics, World History, Party History and Party Construction, and International Studies; the disciplines with the lowest increases included Chinese Literature, Foreign Literature, Law, Applied Economics, Theoretical Economics, and Philosophy.

There have been two relatively large increases in the past 10 years: first, from 2011 to 2013, during which 23 disciplines showed a general increase, with the total number of projects increasing by 32.88%, and the growth rates in 2012 and 2013 were 14.50% and 16.06% respectively; second, from 2016 to 2017, the total number of projects increased by 9.50%, with the fastest-growing disciplines being This growth is closely related to the "May 17 Speech", in which General Secretary Xi Jinping put forward the requirement of accelerating the construction of philosophy and social sciences with Chinese characteristics in his important speech at the Symposium on the Work of Philosophy and Social Sciences. Apart from the above two periods, the number of projects in other years remained basically stable, with slight increases in some disciplines.

The increase in the number of projects under the National Social Science Foundation reflects both the increasing importance attached by the Party and the State to the humanities and social sciences, and the growing demand for and reliance on humanities and social sciences research for China's economic and social development.

3.2. Data Processing Results

Based on web crawler technology and web text mining methods, this paper crawled the text information of the selected topics of the National Social Science Foundation projects, and proposed an LDA strategy for text analysis based on data cleaning and pre-processing operations using python, carried out topic clustering and extraction, and used the software to perform word cutting for each topic of the "National Social Science Foundation" projects. The software was used to explore the cut words of each theme of the "National Social Science Foundation" project. The clustering results of each theme revealed that the high-frequency keywords of the National Social Science Foundation are "China, mechanism, path, era, village, culture and community". Among the annual projects, the themes of "digital, rural, risk, culture and ethnicity" also appear more frequently, while the more prominent themes in the youth projects are "digital, rural, policy, community and culture".

In the field of applied economics, three research frontiers were identified, namely

1. A New Development Pattern of "Double Circles", Both Domestic and International

The new international and domestic development pattern is a long-term strategic deployment by China in light of the global situation. This plan can break the current development dilemma and open up a new pattern of economic development. In the post-epidemic era, the global economic situation is not optimistic. China is facing a loss of human capital and in order to enhance its core competitiveness in global trade, it must make corresponding adjustments and deployments to its development pattern in order to cope with the global economic situation and contribute to healthy and sustainable development. At present, research on the new development pattern of domestic and international "double-loop" mainly

focuses on: (1) how to build a new development pattern of "double-loop" from the perspective of domestic demand-led global value chains; (2) how to promote the new development pattern of "double-loop"; and (3) how to promote the new development pattern of "double-loop". "(2) the capacity base, capacity structure and promotion mechanism for the high-quality economic development in the 14th Five-Year Plan period; (3) the concept, characteristics, development difficulties and realization path of the new development pattern; (4) the theoretical logic and endogenous dynamics of the new development pattern of domestic and international double-cycle.

2.Revitalisation of the Countryside

Accelerating the revitalization of the countryside is a major strategic decision as China's development enters a period of historical convergence between the two centuries of struggle. This not only creates extremely favourable conditions for fundamentally solving the "three rural issues", but also helps to completely break the urban-rural dichotomy formed during China's long-term development and the series of problems derived from it. At the same time, the comprehensive implementation of the rural revitalization strategy is also the key and necessary path on the historical journey to achieve common prosperity. At present, selected research topics on rural revitalization mainly focus on: (1) the importance and motivation of the comprehensive implementation of the rural revitalization strategy; (2) theoretical analysis and practical exploration of rural revitalization and common prosperity; (3) the cultivation and construction methods of farmers' subjectivity; (3) the progress, problems and suggestions of consolidating and expanding the achievements of poverty eradication and rural revitalization; (4) the logic and mechanism of rural revitalization and the return of rural labour. (4) The logic of rural revitalization and the return of rural labor.

3. Digital Governance

Digital governance is an important symbol of modern governance, and the digitization of governance provides the technical means and historical opportunities for the improvement of governance capacity. In the process of digital transformation, both the subject of governance and the way of governance have undergone significant changes, and it is necessary to keep pace with the times and grasp the characteristics of governance, i.e. to reflect the status of the people as the subject of governance and to bridge the digital divide between urban and rural areas. At present, research on digital governance focuses on: (1) the components and theoretical foundations of digital governance; (2) the risks arising from digital governance and their prevention; (3) the bottlenecks and countermeasures for the high-quality development of the digital economy; and (4) digital governance at the government level.

4. Conclusion

It is of great research and practical significance how to quickly and accurately extract the frontiers and hotspots of contemporary research from the vast number of National Social Science Foundation projects, so that future research actors can have a comprehensive understanding of the research results and unresolved issues in related fields within a short period of time. In this study, a combination of LDA and Python is used in text mining. Firstly, Python was used to filter and reduce the dimensionality of the text matrix, and then the LDA topic model was used to extract the topics of each cluster. This strategy effectively reduced the influence of invalid high-frequency words on the topic analysis and improved the topic extraction accuracy. This strategy effectively reduces the influence of invalid high-frequency words on the topic analysis and improves the accuracy of topic extraction. At the same time, a topic clustering analysis is also conducted, and the results can clearly reflect the frequency of occurrence of each topic and related research focus, which not only provides a new idea for research text mining and clustering, but also effectively assists future generations to conduct more efficient related research.

Overall, the NSSF international research projects reflect the overall situation of the current mainstream research in a relatively adequate manner. Through observation and analysis, it is found that the research is generally focused on the keywords "China, mechanism, path, era, village, culture and community". Through observation and analysis of clustering, three major high-frequency modules of applied economics research are summarized, namely, the new development pattern of domestic and international "double cycle", rural revitalization and digital governance research.

Acknowledgement

[Fund Project] This project was funded by the 2022 Undergraduate Research Innovation Fund of the School of Management Science and Engineering, Anhui University of Finance and Economics. (Project number: XSKY22031ZD).

References

[1] Research evaluation in American universities and its reference [J]. Management Observation, Lu Yiyi, Guo Shengwei, 2016 (21).

[2] Gross PF.A critical review of some basic considerations in post-secondary education evaluation [J].Policy Sciences. 1973.4(02): 171-195.

[3] An exploration of the law of development of scientific and technological originality [J]. China Science Foundation, Feng Yueqiang, Qi Wei, 2007(1):14-16.

[4] Research project fund management: a comparative study between China and Britain [J]. Research Management, Gu Quan, 2012, 33(1): 120-126.

[5] Zhou P, Zhang M, Guo S-W. Historical evolution, current situation and countermeasure analysis of scientific research evaluation in China [J]. Management Observation, 2016(32): 173-176.

[6] Analysis of humanities and social science research results in Shaanxi Province[J]. Li Yu,Wang Miao. Intelligence exploration. 2015 (05).

[7] A bibliometric analysis of humanities and social science research in the Chinese Arctic--statistics based on CSSCI journals[J]. Wang Chenguang. Journal of Ocean University of China (Social Science Edition). 2017(02).

[8] A study on the econometric analysis of ethnographic projects of the National Social Science Foundation in the past ten years[J]. Wang Xiaoxia. Comparative study on cultural innovation. 2020(35).

[9] Research progress of Marxist theory disciplines in the western region from the perspective of National Social Science Foundation projects--an analysis of data based on the National Social Science Foundation projects from 2009-2018[J]. Ma

Cunyong, Wang Yongbin. Journal of Lanzhou Jiaotong University. 2020(03).

[10] Analysis report of the National Social Science Foundation of China for Religion from 2008 to 2019[J]. Pei Zhenwei. World Religions.