Rural Revitalization from the Perspective of Data Science

Zhongzheng Sun, Xiaohui Zou and Ruhua Cao

July 25, 2024
Tenth International Conference on Data Science (ICDS 2023)

Rural Revitalization from the Perspective of Data Science

Zhongzheng Sun a, Xiaohui Zou b,c,d *, Ruhua Cao a *

a Institute of Finance, China Academy of Social Sciences, Beijing, 100000, China
b Interdisciplinary knowledge modeling research group, Peking University, No. 5 Summer Palace Road, Haidian District, 100871, China
c SINO-US Searle Research Center, Room 501, Building 1, No. 100 Renshan Road, Hengqin, Guangdong 519000, China
d Institute of Intelligent Engineering and Math., Liaoning Technology University, Fuxin, 123000, China

Abstract

This paper aims at rural revitalization from the perspective of data science. Its method is to adopt three groups of comparative research approaches: man-machine mutual assistance and interpersonal mutual assistance, including personal self-help, and the third method is as follows: First, point out the problems: with the acceleration of urbanization, the factors of production in rural areas in China are greatly drained, the gap between urban and rural development is widening, and the dual economic structure between urban and rural areas is becoming more and more serious. The capitalist mode of production not only harms the interests of farmers, but also makes the workers in cities lose their self-value. Further, explore the way out: the arrival of the digital age has promoted the development of rural areas, guaranteed the interests of farmers and rebuilt the spiritual home of China people. Finally, grasp the key point: analyze the difficulties faced by rural development. At present, there are many difficulties in rural areas, such as resource outflow, low labor productivity of agricultural industry, unclear ownership confirmation of data as a production factor, rigid rural construction structure, etc., and there is a fierce collision between eastern farming civilization and western urban civilization behind it. The result is that, it is proposed to use the means of data value to promote the flow of factors into rural areas and tap effective demand by helping non-commercial finance of primary production. Its significance lies in: it is clear that rural revitalization is currently facing a new era, a new track and a new journey. In the new era, the Chinese nation is experiencing a great rejuvenation, and Chinese civilization will develop new connotations under the collision with western civilization. On the new track, agriculture should seize the digital technology and lead China to take the lead in the fourth industrial revolution. On the new journey, we will build a socialist modern country in an all-round way and embark on a unique Chinese modernization road.

© 2023 The Authors. Published by ELSEVIER B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the Tenth International Conference on Data Science

Keywords: rural revitalization; from the perspective of data science; man-machine mutual assistance; interpersonal mutual assistance; the dual economic structure; digital technology.
1. Main text

This paper aims at rural revitalization from the perspective of data science. So for that we should further explore the innovative practice of smart education guided by the theory of smart system studied as Rongzhi xu, that is, the research on man-machine cooperation assisted by using the STEAM for big data with AI, and for that the application of the man-machine cooperation based on the results show that in the era of big data and artificial intelligence, not only adopts three groups of control studies closely related but harmonious and different. Among them, it includes not only the traditional rule-based expert knowledge system, but also big data processing based on statistics, machine learning and deep learning, a new generation of AI and even primary general artificial intelligence, such as GPT and LLM.[1-8]

Three groups of closely related but harmonious but different places as the nomenclatures are the formal expression tools of the foundation of this smart system studied for the Research on the Mode as e.g.:  

### Nomenclature

A. 2.1. The first group of human-computer dialogue results（Xiaohui Zou and ChatGPT4）;  
B. 2.2. The second group of man-machine conversation results（Xiaohui Zou and Chatbot by JKChat）;  
C. 2.3. The third group is the self-questioning and self-answering（Sun with Cao in Chinese and Zou in English）

2. Methods

Its method is to adopt three groups of comparative research approaches: man-machine mutual assistance and interpersonal mutual assistance, including personal self-help, and the third method is as follows: First, point out the problems: with the acceleration of urbanization, the factors of production in rural areas in China are greatly drained, the gap between urban and rural development is widening, and the dual economic structure between urban and rural areas is becoming more and more serious. The capitalist mode of production not only harms the interests of farmers, but also makes the workers in cities lose their self-value. Further, explore the way out: the arrival of the digital age has promoted the development of rural areas, guaranteed the interests of farmers and rebuilt the spiritual home of China people. Finally, grasp the key point: analyze the difficulties faced by rural development. At present, there are many difficulties in rural areas, such as resource outflow, low labor productivity of agricultural industry, unclear ownership confirmation of data as a production factor, rigid rural construction structure, etc., and there is a fierce collision between eastern farming civilization and western urban civilization behind these difficulties.

What is about “Rural Revitalization from the Perspective of Data Science”? The responses of the two software AI systems are completely different. This has caused us to think: Is it just the difference between the responses of the human-computer interaction software AI system designed based on API in the foreground, or the difference between its LLM in the background? This is not only a problem for data science (mainly in form) but also for smart system (including form and content, its existence principle and construction method). The following is the examples of how two software AI systems face this research topic:

2.1. The first group of human-computer dialogue results（Xiaohui Zou and ChatGPT4 by Time-saving technology）

**People ask question:** What is about “Rural Revitalization from the Perspective of Data Science”?

**Machine answer** (artificial intelligence system gives generalized experience): From the perspective of data science, we can use technologies such as data analysis and machine learning to help decision makers better understand and promote rural revitalization strategies. The following are some possible data science methods and applications:

Data collection and analysis: collect and sort out all kinds of rural data, such as demographic data, land use data, economic data, etc. Through the analysis of these data, we can understand the present situation, problems and potential of the countryside.
Prediction and model building: using historical data and machine learning algorithm to build a prediction model to predict the trend of rural development and possible influencing factors. This is helpful to formulate reasonable rural revitalization strategies and plans.

Spatial analysis and optimization: Using geographic information system (GIS) and spatial analysis technology, the rural spatial layout, resource utilization and transportation network are analyzed and optimized. This can help decision makers to plan the development direction and resource allocation of rural areas reasonably.

Social network analysis: By analyzing the social network relationship among rural residents, we can understand the organizational structure and information dissemination mode of the community. This will help to discover potential cooperation opportunities and community dynamics and promote rural development.

Data-driven decision support: based on the results and models of data analysis, it provides decision support and recommendation for decision makers. This can help them to formulate more scientific and effective rural revitalization policies and measures.

In a word, data science plays an important role in rural revitalization. By applying data science methods and technologies, we can better understand the current situation and needs of rural areas, formulate scientific development strategies, and provide decision support to promote the sustainable development of rural revitalization.

2.2. The second group of man-machine conversation results (Xiaohui Zou and Chatbot by JKTChat)

**People ask question**: What is about “Rural Revitalization from the Perspective of Data Science”?

**Machine answer** (artificial intelligence system gives generalized experience): Rural revitalization is an important issue facing China at present, which needs to be analyzed and explored from the perspective of data science. The following are several aspects of rural revitalization from the perspective of data science:

1. Data-driven rural industry development

   Industrial development is an important direction in rural revitalization. By collecting and analyzing the industrial data in rural areas, we can understand the local industrial development and market demand, so as to formulate the corresponding industrial revitalization plan. For example, by collecting local data on agriculture, industry and tourism, we can analyze the demand and development of different industries in order to make more targeted revitalization plans.

2. Data-driven rural infrastructure construction

   Infrastructure construction is also an important aspect in rural revitalization. By collecting and analyzing the data of transportation, water conservancy and electric power in rural areas, we can understand the local infrastructure and make corresponding revitalization plans. For example, by collecting local power and traffic data, we can analyze the usage of infrastructure so as to make corresponding improvements.

3. Data-driven rural social governance

   Rural social governance is also an important aspect in rural revitalization. By collecting and analyzing the data of population, crime and traffic in rural areas, we can understand the local social governance and make corresponding revitalization plans. For example, by collecting local population data, we can analyze the local population structure in order to make corresponding social governance plans.

4. Data-driven rural ecological environment protection

   In rural revitalization, ecological environment protection is also an important aspect. By collecting and analyzing the data of environment, transportation and agriculture in rural areas, we can understand the local ecological environment protection and make corresponding revitalization plans. For example, by collecting local traffic data, we can analyze the impact of traffic on the environment in order to make corresponding environmental protection plans.

   Driven by data, it can help rural revitalization to be better realized.
Based on the experimental results of the above two human-computer dialogues, the following basic analysis can be made: firstly, it can be seen that the general strategies of the two design and production teams are completely different. (as to how much of his answer meets the meaning of the question, or satisfies the questioner or can be recognized to what extent? That is a problem that anyone will encounter in the process of further active exploration). This is the most fundamental difference between the responses of the human-computer interaction HCI software AI system based on API design in the foreground. Further, it can be seen that there are at least two situations in which two foreground softwares use background LLM, either directly adopting the GPT and LLM of OpenAI, or using the GPT independently copied by open source software or using other LLM, but as far as we know, at least they using the corresponding LLM. This is not only a preliminary analysis and judgment from data science (mainly from the form) but also from smart system science (from the form and content, the principle of its existence and the method of its construction). Among them, the problems that need further study. In-depth study can be made from interpersonal, man-machine and inter-machine (between machines) separately or even in combination. Let's compare the above two examples of human-computer interaction and even cooperation and mutual assistance to see how individuals and interpersonal people think, understand and deal with the same topic or theme:

2.3. The third group is the self-questioning and self-answering (Sun with Cao in Chinese and Zou in English)

**Question:** What is about “Rural Revitalization from the Perspective of Data Science”?

**Answer:**

On October 18, 2017, the report of the 19th National Congress of the Communist Party of China put forward the strategy of rural revitalization, and it has been six years since then. Under the leadership of the Communist Party of China and the unremitting efforts of the people, many gratifying achievements have been made in rural revitalization. Chinese government introduced a lot of supporting policies and measures for the development of rural areas, greatly improving the efficiency of rural productivity.

At present, while people all over the world need to develop together, cities and villages need to develop together. We must recognize the difficulties in urban and rural development, think about the roots of urban-rural contradictions, and finally embark on a Socialism with Chinese characteristics road.

**2.3.1. First, the question raised**

With the rapid progress of urbanization, people have noticed that the dual urban-rural economic structure is becoming increasingly severe, and cities are absorbing resources from rural areas in a desperate way, shaking the foundation of rural revitalization. Due to many reasons, such as policy inclination, capital alienation, ideological infiltration, etc., farmers can't get the same income from rural land as urban land at present, so a large number of rural migrant workers go to cities to work, and the factors of production such as labor force, capital, technology and even land continue to flow from rural areas to cities. However, the increase of factor input has already entered the range of diminishing marginal returns in cities, and urban workers work hard but do not earn much disposable income, accompanied by a high unemployment rate. In rural areas, farmers, as the main body of rural revitalization, cannot enjoy the same social welfare as urban residents. In the end, whether urban or rural residents become more and more toilsome and lost their spiritual homes.

Marx's political economy pointed out that the capitalist mode of production is a mode of production whose nature is determined by the relationship between capital and wage labor. Socialized production greatly improves labor
productivity, but the factors of production are privately owned by capitalists. The current urban development is the product of capital operation.

Rural revitalization can no longer take the road of urban development. As the primary industry, agriculture is the base of all other industry, and its special position determines that agricultural production must be stable and reflect the value of labor. rural areas have their own unique development model, which means that villages should maintain their diversity.

2.3.2. Second, the plight of rural development

(A) Analysis of the contradiction between urban and rural residents

In the process of urbanization, the most obvious difference between urban and rural areas is the population gap and the income gap.

The population gap reflects the profound changes that urbanization has brought to the urban-rural pattern for decades. At present, with the acceleration of urbanization, the rural population is facing the dual pressures of high outflow rate and low birth rate.

The first pressure comes from the high outflow rate of rural population. According to the data of urban and rural population changes, by the end of 2022, Chinese urban resident population reached 920.71 million, an increase of 6.46 million over 2021; The resident population in rural areas was 491.04 million, a decrease of 7.31 million. China has gradually changed from a large rural population to a large urban population. However, urbanization do not change the dual structure of urban and rural areas, but lead to a large loss of population in rural area.

The second pressure comes from the low birth rate of rural population. At the end of 2022, Chinese population was 1,411.75 million, and the natural population growth rate was -0.60‰, which means that the population of China experienced negative growth for the first time in more than 60 years. Population decline is the result of many factors, one of the most important factors is the continuous decline of women of childbearing age. It can be predicted that in the next ten to twenty years, the rural population will decline like a cliff.

The income gap is also the contradiction between urban and rural residents. The reason for the loss of rural population is that the income from agriculture is not enough for farmers to survive, and the widening income gap between urban and rural residents urges people in rural areas to work in cities.

The term "rural migrant workers in cities" refers to workers whose household registration is still in rural areas and who have been engaged in non-agricultural industries in cities. In 2022, the total number of migrant workers in China was 295.62 million, an increase of 3.11 million or 1.1% over the previous year. In the process of urbanization, migrant workers have become neither rural residents nor urban residents in the traditional sense. This special group faces many problems in urban life, such as income insecurity, poor living conditions and lack of social security. What's frustrating is that radical urbanization construction will undoubtedly further widen the income gap between urban and rural residents.

(B) Analysis of the contradiction between developed urban industries and backward agricultural production

The contradiction between industry and agriculture can be divided into three levels from shallow to deep, namely, the output value gap, the financial supply gap and the deviation from traditional farming civilization.

The most obvious is the huge output gap between urban and rural areas. Urban industry represented by real estate has a complete industrial system and an easy-to-copy development model. Real estate is the foundation of urban industry, and it can promote a series of related industries. In contrast, agriculture has a huge gap as a systematic industry. Rural land is based on cultivation. Under the premise of controlled grain prices and slow growth of grain output, there is a lack of related industries in agriculture at present, so the output value of rural agricultural system would be far lower than that of urban industrial system.

Since 1978, the share of the primary industry in the gross national income has always been far lower than that of the secondary industry and the tertiary industry, and the average growth rate of the primary industry is only 4.4%, far lower than the average growth rate of the secondary industry and the tertiary industry.
The output value of rural land lagging far behind that of urban land, directly leads to the huge gap of financial supply between urban and rural areas. Commercial financial institutions are keen to tilt financial resources towards urban development in pursuit of higher expected returns. Urban land has a higher output value. After asset mortgage, people can easily obtain sufficient funds for reinvestment, which in turn drives up the output value of urban land again. However, the supply of financial resources in rural areas is relatively deficient. Due to many factors, such as low output value, inconvenient land use right transaction, and the influence of natural environment such as weather and climate on agricultural output, rural areas have been relatively lacking in financial support. As a result, the imbalance in funding supply will widen the gap between rural and urban areas.

Buried under the gap between industrial value and financial supply is the fierce collision between eastern farming civilization and western urban civilization. Chinese people have special feelings for the land, which is the direct embodiment of Chinese farming civilization. Land is not only a basic factor of production, but also a spiritual home for each Chinese person.

But western urban civilization is now invading Chinese opinion of value, and the idea of pursuing capital expansion and maximizing personal interests has already rooted in people's minds. Urban land has seriously deviated from its value under the drive of capital, and workers have gradually been alienated into tools of production and commodities by capital. In the end, people price their value with a number. People would lose their direction and enter into a low desire society at the same time.

(C) Analysis of the confirmation of agricultural big data

The development of digital economy has greatly changed the appearance of rural areas, and digital technology has been deeply integrated with rural development. With the continuous improvement of rural digital infrastructure, we can see the blueprint for the application of agriculture big data, which is the key point for agriculture upgrading.

As a newly developing factor of production, data is attracting more and more attention. The input of data factors can significantly increase labor productivity, and tap the demand that was originally difficult to detect. There are many characteristics that other production factors do not have. Jones & Tonetti(2020) established a theoretical framework of data based on economic principles, pointing out that the use of data is non-competitive. If the private sector holds the data property right, it will hoard data for fear of creative destruction. However, if consumers master the ownership of data, they will be able to achieve near-optimal resource allocation.

People in rural area are the creators and consumers of agricultural data, and they should be given the property rights of agricultural data. But actually farmers and collectives’ voice in determining data ownership is weak. The processing flow of big data can be summarized into four steps, namely, collection, import and pretreatment, statistics and analysis, and mining. Taking agriculture data as an example, from the first step of collecting data, once farmers lease out the land, the data would be collected and occupied by large agriculture groups supported by private capital. In fact, private capital is the actual user and beneficiary of data.

2.3.3. Third, policy recommendations

(A) monetary and financial support for agriculture

Rural revitalization can not do without financial support, but the direct intervention of traditional commercial finance will inevitably lead to the capital concentrate too much on the yield to maturity but ignoring the need of rural development. Rural revitalization requires the support of non-commercial finance and innovation in the central bank's monetary policy
According to the asset side of the central bank's balance sheet, the supply of money is essentially anchored by national credit and foreign exchange, and the profit-seeking nature of capital leads to the deep binding of national credit with real estate, finance and other high net worth assets. On the one hand, real estate indeed played a pillar role in economic development, but the disorderly expansion of cities also left a huge hidden risk. Once there is a crisis in real estate, such as liquidity risk or duration risk, banks would immediately face the risk of asset impairment. On the other hand, as for foreign exchange reserves, we can see that the recent aggressive interest rate increase policy of the Federal Reserve has led to the bankruptcy of Silicon Valley Bank, Credit Suisse and other large commercial banks. Chinese central bank, which also holds a large amount of US debt, has to be alert to its risks.

When the real estate is weak and the USD credit is in jeopardy, the central bank needs more stable assets. On the basis of the existing assets of the central bank, adding commodities that can directly reflect the production value of rural labor, such as grain, will make the bank's asset value more stable and reduce the possibility of systemic risks.

Then considering the liabilities of the central bank, it is important to ensure that the liquidity provided by the central bank can be truly invested in the real economy and ultimately converted into labor income. In China, e-CNY may become the key to this problem. E-CNY is the digital Fiat money issued by Chinese central bank, which is convenient, anonymous and traceable. And e-CNY can accelerate the velocity of money, increase the production efficiency, and reduce fund misappropriation and corruption. Eventually, the improvement of overall social welfare will increase the multiplier effect of macroeconomics.

(B) to promote the integration of agricultural industries, improve the "double chain" construction.

In order to increase the output value of agriculture and its related industrial systems so that they can compete with urban industries, it is necessary to promote industrial integration and improve the "double chain" construction of industrial chain and supply chain. Imamura Naraomi, an honorary professor at Tokyo University in Japan, put forward the concept of "the sixth industry", which gradually expanded the original single primary industry into a comprehensive industry through diversified operations, and created more added value in the process of supporting the secondary industry and the tertiary industry. Nowadays, most of the tertiary economy is based on the secondary economy, but almost divorced from the primary economy. In the competition between the primary industry and the secondary industry, investors tend to prefer the secondary industry, which is a stereotype. Supported by data science, new agriculture can fully utilize spatial and environmental resources, achieving benefits several times higher than traditional agriculture. We believe that agriculture can serve as a source of vitality for a country's economic growth.

Agriculture industrial integration will put forward higher requirements for industrial chain and supply chain, both of which need the support of data science. At present, there are some problems in the development of industrial chain and supply chain of smart agriculture in China, such as high risk, low output and weak infrastructure. People need to utilize data science to promote the flow of information between industries, improve infrastructure, and thereby reduce costs and increase benefits in the production and supply chains.

(3) Improve the ownership system of data factors and increase human resource investment in data science.

With the advent of the digital economy era, the labor productivity of agriculture upgraded and the overall appearance of the countryside improved, but the most important thing is to let people in rural area enjoy the interest of industrial upgrading. In the digital age, big data is likely to bring a new industrial revolution, and it is necessary to stand at the height of Marxist political economy and promote the determination of ownership of data production factor. It will directly affect whether labor is the master of surplus value or continue to be exploited by capital. In China, the people are the masters of the country and society, and the government will protect the interests of the widest range of people. It is necessary to ensure the position of public ownership in data factor, strengthen the legislation and supervision related to the application of data, and let the whole society benefit from the industrial upgrading brought by data science. In addition, it is significant to increase human resource investment in data science, especially for officials and builders in rural areas. For a long time, as a region where capital and talents flow out in one direction, villagers or village collectives have always faced problems such as not knowing, not using well and not paying attention to the data science. To reverse people's cognition, the government's rural departments need to increase their efforts in publicizing data science. Also it is necessary for rural areas to actively cooperate with universities or research institutes to attract talents in data science to rural areas, promoting the utilization rate of agricultural technology.
3. Results and Conclusion

The result is that, not only the responses of the two software AI systems are completely different, but also the comparative advantages of the three research approaches are further discovered and that, it is proposed to use the means of data value to promote the flow of factors into rural areas and tap effective demand by helping non-commercial finance of primary production.

Its significance lies in the fact that this study has caused us to think: Looking back at the examples of how two software AI systems face the research topic of human beings, it is of great significance for many years of experimental results recorded in Chinese, English reporting and bilingual thinking. What needs to be emphasized is the third group, which is not so much the practice of interpersonal interaction as an attempt to improve the existing practice in the era of man-machine mutual assistance, and its significance lies in it is clear that rural revitalization is currently facing a new era, a new track and a new journey. In the new era, the Chinese nation is experiencing a great rejuvenation, and Chinese civilization will develop new connotations under the collision with western civilization. On the new track, agriculture should seize the digital technology and lead China to take the lead in the fourth industrial revolution. On the new journey, we will build a socialist modern country in an all-round way and embark on a unique Chinese modernization road.

Acknowledgements

Thank Professor Jing He from Oxford University for whom inviting us to attend the meeting. Thanks to the tow teams of both the ChatGPT4 and the JKTchat.

References