

Research on the Relationship Between Regional Economic Sustainable Development and Human Land Coordination Based on Data Mining under the Background of Big Data

Yu Xin^{1,*}

¹ Xi'an International University
Xi'an, Shanxi, China

Abstract — In this paper, the author studies the relationship between regional economic sustainable development and human land coordination based on data mining under the background of big data. In this research work we focused mainly on the precise mobility profile building through trajectory and behavioral pattern mining using the GSM CGI Cell-ID, where all the concerned issues like precise spatial extraction, stay points detection and mobility profiling are addressed properly through the proposed framework. The proposed framework utilized both spatial term and semantic information for mobility profile building which is not addressed in any of previous related work, so makes it suitable for any LBS due to of its novel and generic nature. The result shows that by using data mining under the background of big data, we can get the quantitative description of the relationship between regional economic sustainable development and human land coordination.

Keywords - regional economic sustainable development; human land coordination; data mining; nonlinear mapping approximation performance

I. INTRODUCTION

Sustainable economic development strategy is the inherent requirement of human economic and social development, and regional resources and environment is the basis of regional economic development. For a long time, China's regional economic rapid growth is largely based on high consumption of resources and energy, the traditional model of development inevitably resulted in environmental pollution. The deteriorating ecological environment has become the constraints of regional economic sustainable growth. Exploring regional economic development path under environmental constraints has become the key issues of speeding up economic development pattern and achieving sound and rapid development of regional economy.

Factor agglomeration is a unique mode of resource allocation in the context of economic globalization, and has a huge role on promoting local economic growth, but in the process of regional economic growth should pay attention to the coordination of the economy and the environment. Only to develop regional innovation capacity, promote regional economic growth model to change from material elements-driven into the innovation-driven, until achieve regional economic development. According to the breakthrough role of regional innovation capability on resource and environmental constraints on the process of regional economic development, regional innovation capacity as mediating variables, we along timing evolution path of factor agglomeration to regional innovation capability, to regional total factor productivity growth, analyze influence mechanism of factor agglomeration on regional innovation capability, dynamic mechanism of regional innovation

capacity on total factor productivity, revealing underlying mechanisms of factor agglomeration on promoting coordinated and sustainable development of regional economy under environmental constraints from the dimension of time-space.

Factor agglomeration is that various factors of production (labor, physical capital and human capital relying on the economic activities of enterprises, by a certain structure gather in a given region under the guidance of internal expansion and spatial concentration of more businesses, and through effect of factor agglomeration enhance regional innovation ability, then promote regional economic development. Regional economic development is not only the quantitative growth, but also focus on the growth quality. Regional economic development is considered from three aspects: Firstly, the expansion of the regional economic scale and improvement of production efficiency; Secondly, the optimization of regional spatial structure.

Focusing on the application of green, clean technologies to the regional economic development, not damage the ecological environment. Regional economic development must rely on regional innovation capability. Regional innovation capacity refers to capabilities of improving the stock of knowledge and skill levels, innovation subjects of a certain area directly or indirectly involving in the production of knowledge or technology innovation, through integration and convergence of innovative resources. Regional innovation capability was analyzes from the four aspects of knowledge creation, knowledge acquisition, corporate innovation and the innovation environment.

The characteristics and advantages of the ecological environment through the mobility of factors and location changes determinants factor agglomeration. Factor

agglomeration has brought huge economic benefits to gathering area, also brought negative effects to the ecological environment. Regional Innovation capability is the fundamental way to achieve win-win of regional economic development and environmental protection. Taking carbon dioxide emissions as an alternative to variable environmental pollution, this article measures carbon footprint of energy consumption, and measures density, location quotient and spatial autocorrelation of labor, physical capital and human capital, and found that labor, physical capital and human capital are highly concentrated and positive spatial correlated. This paper uses the Divisia index decomposition method to analyze the scale effect, technological progress effect, regional industrial structure effects and spatial structure effects of regional economic growth on environmental pollution emissions. The results showed that unreasonable of the scale and regional industrial structure is the main reason for environmental pollution, regional spatial structure play an adverse impact on the improvement of the environment quality; technical effects play a positive role on promoting the improvement of the environment quality.

II. THE FRAMEWORK OF BIG DATA AND BEHAVIOR ANALYSIS ARCHITECTURE

Nowadays, big data has become an important direction of development of modern information technology, and sharing and analysis of big data would not only bring immeasurable economic value, but also play a significant role in promoting

the development of society. Big Data-as-a-Service (BDaaS) is a new data resource usage pattern and a new form of service economy, by encapsulating heterogeneous data; it can provide ubiquitous service consumers, standardization, and on-demand services, including search, analysis or visualization.

In behavior analysis area, it still faces four challenges:

- 1) There is no standardized, user experience based BDaaS architecture which can shield the complexity of data sources and operations;
- 2) The lack of generic unstructured data model which reflects user behavior characteristic, made BDaaS for unstructured data difficult to build;
- 3) Existing data model follows the Web services model, however, so far, holistic BDaaS service model with the characteristics of big data has not yet appeared;
- 4) There is no appropriate solution in providing data retrieval, analysis and visualization services and optimizing service capacity. However basic network architecture of GSM network is same for every network operator as shown in Figure 1.

Today, with the advancement in information technology and the development of the economic, internet has gotten a more and more important position in our life. The rapid expansion of Internet, network traffic, network user and the number of host computers increases with the exponential growth.

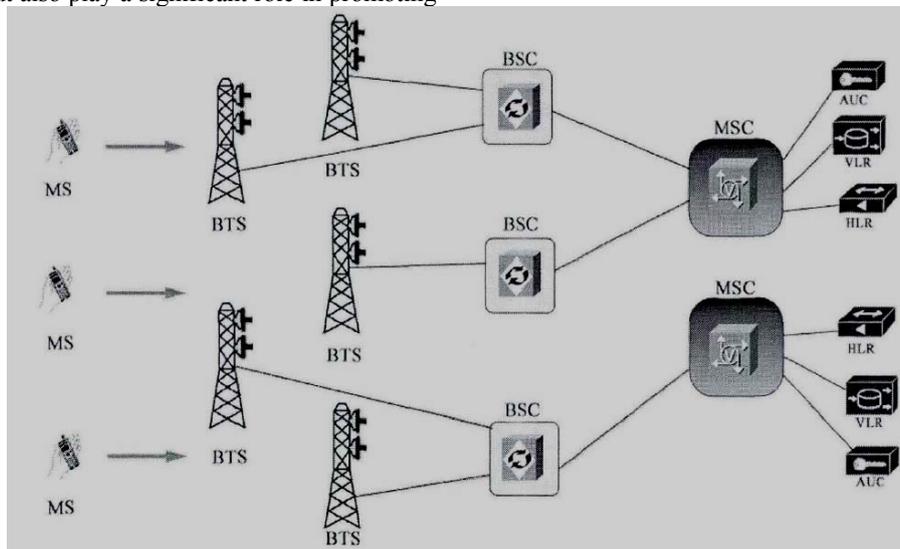


Figure 1. The model and architecture.

With the development of mobile devices, embedded systems and sensor networks, the new Internet element, Internet-scale will continue to grow in a long time. Internet applications have developed from simple traditional applications to real-time multimedia applications. Nowadays Internet applications have the new characteristics that sharing the resource and collaboration. However, the important reason for Internet reaching today's position is that

the rapid development of Internet services, network application, diversified types of business and personal trends. In a closed or semi-open network which, the security is assured; this paper uses data mining algorithms of Bayesian incremental learning to classify user data for obtaining the characteristics of user behavior and data model then using Bayesian incremental learning algorithm to optimize the model to orientate consumer behavior accurately. In addition,

the user behavior analysis system can connect directly with operator's business systems. More specifically, the user behavior analysis system can help to find more potential

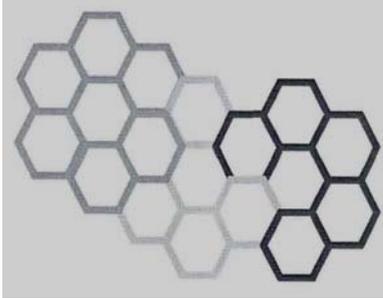


Figure 2. The network topology.

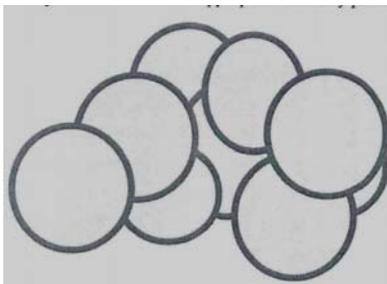


Figure 2. The network topology: Actual bubble shape.

III. DATA MINING ALGORITHM

The main effect of data mining algorithm is to extract the required information from a large amount of data, including the structural data, semi-structured, and unstructured data sources, such as audio, video, data, for data algorithm. This algorithm must have model, and first search algorithm. Currently the common data mining algorithms are mainly the decision tree method, bionic global optimization of genetic algorithm and neural network, statistical analysis and row exclusive counterexample method, etc. In order to improve the effect of data mining process effectively, a detailed research should be applied to the cloud computing method. In this way, more effective implicit knowledge in the mass data information can be discovered in order to improve the effect of application of information data.

Association rules found a relationship between things and other transactions or interdependence. Assuming that $I = \{i_1, i_2, \dots, i_m\}$ is a collection and the related data task D is a collection of database transactions, in which each transaction T is a collection, and making $T \subseteq I$. Every transaction has an identifier TD . Assuming that A is a set of items, $A \subseteq T$. Association rules are the containing type of $A \Rightarrow B$, among them, $A \subset I$, $B \subset I$ and $A \cap B = \Phi$. The rules $A \Rightarrow B$ is in the transaction which sets up with support s , s is the percentage for the transaction contains $A \cup B$ in the D .

users in the way of advertising. We analyze the 3G networks operators that are meeting brand-new business challenges, introducing the system diagram of user behavior analysis.

The LIPI algorithm through scanning data sets of frequency, then finding relevant data and finishing dig, its principle as follows:

Assuming that $I = \{i_1, i_2, \dots, i_n\}$ is a collection, which composed of different characteristics, the characteristics of each item as a constituted set of items. And the item set is not an empty set, but is a subset of the set of I , which can be expressed as (x_1, x_2, \dots, x_m) , every x_k is a term.

The sample variance and sample proportion of variance have established the following relationships:

$$S^{*2} = \frac{S^2}{\mu^2} \tag{1}$$

Proof: by definition 1

$$\begin{aligned} S^{*2} &= \frac{1}{n-1} \sum_{i=1}^n \left(\frac{\alpha_i - \bar{\mu}}{\bar{\mu}^2} \right)^2 \\ &= \frac{1}{(n-1)\bar{\mu}^2} \sum_{i=1}^n (\alpha_i - \bar{\mu})^2 \\ &= \frac{1}{\bar{\mu}^2} S^2 \end{aligned} \tag{2}$$

The algorithm is to propose the original data processing for many times, then use the effective information contained in the original data.

The study of the data mining algorithm has great significance in improving the effect of data processing. User data information needs to be extracted in the huge amounts of data, in order to promote the development of various fields. Cloud computing is a relatively new computing mode, its application in data mining also needs to do some further researches on the existing basis, continuously improving its application efficiency and improving data information of the application efficiency.

The data mainly displays in some aspects, such as by the department of statistics data format, which does not have a uniform requirements. When using VF or SQL or some using TEXT or a variety of other formats, each system has general check, data format, and compatibility.

Complicated statistical data have originated from the enterprises and institutions' direct submitting, or the result of the whole system of internal between different departments. For the lack of effective statistical data storage and management of professional means, it leads to the deep processing of statistical data.

It is known that the statistics business involves all aspects of society including index the great amount of data. Although there are abundant data resources, yet the lack of professional analysis of statistical data at a deeper level and lack of refining and mining tools, leads to the disunion between a large number of accurate data and resource usage and people's growing demand of statistical information.

The decision of government and enterprises and the current statistical work has been simply filled in form, and then submitted, but it is lack of effective means for the subsequent development. The data mining of data warehouse technology can solve the above mentioned problems effectively mainly because it has the following obvious advantages:

(1) Based on data warehouse algorithm, the data time-consuming preconditioning in data mining can be solved. Through the establishment of data warehouse, it can avoid data extraction, cleaning, conversion and loading process every time.

(2) Another feature of the data warehouse is to store data through subject organization, which provides convenience for data mining to choose the appropriate data source. According to different areas, data warehouse is divided into the national economic statistics, social statistics analysis, and enterprise survey, etc. The national economy includes consumption statistics, labor statistics, people's living standard, and statistics, etc.

(3)The data is collected by statistics departments at all levels, all existing in different types of database such as EXCEL, Fox Pro, etc. Because historical data cannot be stored in the database, many knowledge can't be excavated in the mining database, such as forecasting and application; on the other hand, data warehouse storage management can get the data from the PLTP system and the history of offline business data and the external data sources of heterogeneous distributed, thus it's good for the heterogeneous data and source data to summarize in order to finish the more efficient usage. Except for the requirements of data mining and data warehouse environment, data mining needs to be based on data cube environment and data warehouse technology and hence meet the demands of data mining technology. Therefore, it is necessary for data mining and data warehouse to work together. On the one hand, data mining technology has become a very important application of data warehouse and a relatively independent tool; On the other hand, the data mining technology is an important step of catering the process of data mining, and it improves the efficiency of data mining and ability, and ensures that data mining have the extensiveness and completeness of data source.

IV. EXPERIMENT RESULTS

Physical capital agglomeration shows the characteristics of capital deepening, but capital deepen promoting economic development can be established only under the premise of capital deepen promoting regional innovation capability. Agglomeration of human capital is conducive to the generation of trust and the spread of knowledge, especially tacit knowledge, resulting in knowledge spillovers. By building space knowledge spillover model, analyzes the effects of physical capital investment (capital deepen), agglomeration of human capital on the growth rate of the stock of knowledge, found that the agglomeration of human capital improves the regional absorptive capacity, increasing the knowledge spillover, investment in physical capital embodies the learning by doing effects of knowledge

production, increasing the human capital and physical capital investment of the backward areas helps narrow of the knowledge gap, and the realization of regional economic convergence. This paper makes empirical analysis of agglomeration of material capital and human capital, the intensity of energy footprint and environmental regulation level on regional innovation capability using GWR model. Table 1 shows the sequence database and table 2 shows the CA sequence database.

The growth of the regional environment total factor productivity reflects the quality of regional economic development. The paper constructs a mathematical model to analyze the dynamic mechanism of the regional innovation capability and the reduction of pollution emissions on regional total factor productivity growth. On this basis, estimate and decompose the technical efficiency and total factor productivity growth of Chinese provinces and cities using the SBM directional distance function and the GML index under environmental constraints. The results show that distribution of environmental efficiency in inter-provincial is uneven; technological progress and resulting in increasing returns to scale is dominate to promote environmental total factor productivity growth.

On the basis of controlling of the industrial structure, foreign direct investment, the degree of market variables, using Tobit model to analyze the effect of environmental regulation, regional innovation capacity on environmental efficiency and total factor productivity growth. It was found that regional innovation capability has a significant role on promoting environmental efficiency and total factor growth rate; government environmental regulation is negatively related to environmental total factor productivity of environmental efficiency.

TABLE I SEQUENCE DATABASE

S _{id}	Sequence
S ₁	ACAABC
S ₂	ACACB
S ₃	ACABC
S ₄	ABCAB

TABLE II CA SEQUENCE DATABASE

S _{id}	Sequence
S ₁	ABC
S ₂	BCB
S ₃	BC
S ₄	B

We have implemented the proposed methodology on the dataset which is spatial outlier free. Our methodology is based on overlapping area so we plotted the cell appearance over different locations which show the bubble effect in GSM network as shown in Figure 4. And also shown in Figure 5 where one semantic location is identified by

multiple cells and one cell can appear at multiple semantic locations. The results show that capital deepen of physical capital agglomeration doesn't transform into technology deepening, and not promoting regional innovation capability; human capital agglomeration significantly promotes regional innovation capacity; the greater intensity of energy footprint is, the greater constraints role on the regional innovation capacity is; Environmental Regulation dual effect on regional innovation capability, and current environmental regulation policy is not conducive to improvement of regional innovation capability, while incentives role of the pre-environmental regulation on regional innovation is highlighted to make up, even more than the negative effect of environmental costs.

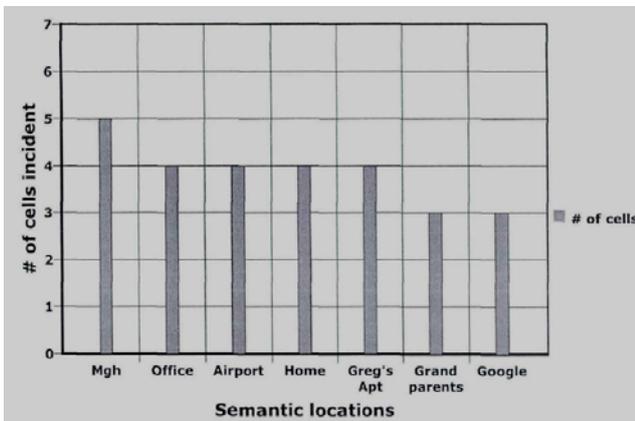


Figure 4. The data mining result on network topology.

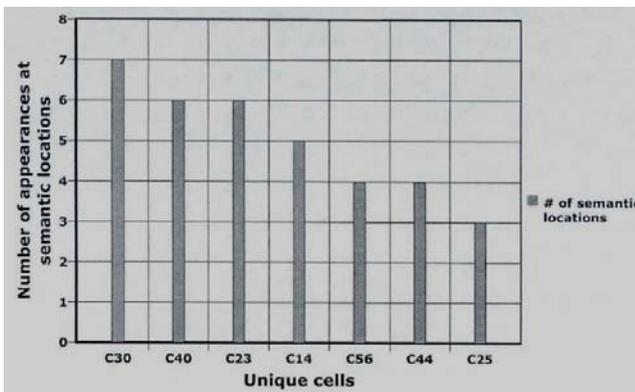


Figure 5. The users' behavior analysis data.

Endogenous mechanisms of factor agglomeration on sustained economic growth under environmental constraints. Constructing endogenous growth model that contains factor agglomeration and environmental quality, and using the Hamilton dynamic optimization model to solve the optimal steady-state growth solution, this paper discusses necessary conditions for sustainable economic development, and the dynamic relationship of agglomeration external level, consumer and environmental preferences, pollution emission density and the rate of economic growth. The results show

that on the equilibrium path of sustainable economic growth, the greater the spatial externality of agglomeration is, the faster the rate of human capital accumulation is, the greater the density index of pollution emissions is, the lower time discount rate is, the greater the preference of environmental quality is, the smaller inter-temporal elasticity of substitution is, then the greater the rate of economic growth is. Using panel fixed effects model to make an empirical analysis of factors affecting regional economic development, the results show that physical capital agglomeration and regional economic growth show the inverted U-shaped relationship, excessive physical capital agglomeration hinders the improvement of labor productivity; human capital, gathering a significant role in promoting labor productivity; an invalid environmental efficiency significantly hinders the growth of labor productivity; only human capital agglomeration by acting on regional innovation capability indirectly promote regional economic growth. The experiment builds a local spillover model to analyze active mechanism of external economies of scale effect of spatial agglomeration, which are demand related and spillover, and external diseconomies of environmental pollution, crowding on regional economic coordinated development. The results showed that demand related effects and spillover effects are gathered force, environmental pollution, crowding and other external diseconomies are dispersion force to maintain balance symmetric. At the same time, spillover effects can also be a force for economic dispersion.

There is "inverted U" relationship between industry agglomeration and economic development. Due to spillover effects, the faster economic growth, the stronger regional external radiation, the more conducive to narrow the regional gap. Regional economic cooperation is conducive to regional knowledge spillover and economic radiation, but also it is an important way to address regional environmental pollution. Build a game model of regional economic cooperation to analyze the utilities allocation problem of regional economic cooperation and the formation of cooperation mechanisms. Local governments participating in cooperation brings incremental of the economic utility and loss of environmental utility, the difference between the two is the net utility of local governments participating in cooperation bringing to the union, allocating utilities of regional economic cooperation according to this proportion of net utility. The sufficient condition of formation of regional economic cooperation mechanisms is that utility of local governments participating in cooperation is greater compared to not participating in cooperation. Complementary of factors is a necessary condition for regional economic cooperation.

V. CONCLUSION

The relationship between regional economic sustainable development and human land coordination based on data mining under the background of big data is researched in this paper. Sustainable economic development strategy is the inherent requirement of human economic and social development, and regional resources and environment is the basis of regional economic development. In this research

work we focused mainly on the precise mobility profile building through trajectory and behavioral pattern mining using the GSM CGI Cell-ID, where all the concerned issues like precise spatial extraction, stay points detection and mobility profiling are addressed properly through the proposed framework. The proposed framework utilized both spatial term and semantic information for mobility profile building which is not addressed in any of previous related work, so makes it suitable for any LBS due to of its novel and generic nature.

For a long time, China's regional economic rapid growth is largely based on high consumption of resources and energy, the traditional model of development inevitably resulted in environmental pollution. The deteriorating ecological environment has become the constraints of regional economic sustainable growth. The experiment conclusion gets that there is "inverted U" relationship between industry agglomeration and economic development. Due to spillover effects, the faster economic growth, the stronger regional external radiation, the more conducive to narrow the regional gap. The result also shows that by using data mining under the background of big data, we can get the quantitative description of the relationship between regional economic sustainable development and human land coordination.

REFERENCES

- [1] D. McDonagha, A. Bruseberg, C. Haslamc, "Visual product evaluation: exploring users' emotional relationships with products". *Applied Ergonomics*, 33, p.p.231-240, 2002.
- [2] Kaikai Chi, Yi-hua Zhu, Xiaohong Jiang, Xianzhong Tian, "Practical throughput analysis for two-hop wireless network coding". *Computer Networks*, pp. 233-256, 2013.
- [3] Luiz Filipe M. Vieira, Mario Gerla, Archan Misra, "Fundamental limits on end-to-end throughput of network coding in multi-rate and multicast wireless networks". *Computer Networks*, pp. 5717-5727, 2013.
- [4] Yang Xu, Xiao yao Xie, Huan guo Zhang, "Modeling and Analysis of Electronic Commerce Protocols Using Colored Petri Nets". *Journal of Software*, pp. 67-78, 2011.
- [5] D.E. Leidner, "Virtual partnerships in support of electronic commerce: the case of TCIS". *Journal of Strategic Information Systems*, pp. 81-93, 1999.
- [6] Jiali Yun, Liping Jing, Jian Yu et al, "A multi-layer text classification framework based on two-level representation model". *Expert Systems with Application*, 39(2), pp. 2035-2046, 2012.
- [7] Changxing Shang, Min Li, Shengzhong Feng, Qingshan Jiang, Jianping Fan, "Feature selection via maximizing global information gain for text classification". *Knowledge-Based Systems*, pp. 54-68, 2013.
- [8] D. McDonagha, A. Bruseberg, C. Haslamc, "Visual product evaluation: exploring users' emotional relationships with products". *Applied Ergonomics*, 33, p.p.231-240, 2002.