# Research on Rural E-commerce Reverse Logistics Network Based on Constrained Clustering Algorithm

Caihui Guo[1,2], T. Khukhmurun [1], Rongxia Wang[2], Lian Dai[2]

[1]*Ulaanbaatar Erdem University, Ulaanbaatar 999097-15141, Mongolia*
[2]*Guangzhou nanyang polytechnic college, Guangdong 510900, China*

**Abstract**

In recent years, the state has been committed to promoting the growth of rural e-business, and has elevated it to the level of accelerating rural economic growth, increasing farmers' income and accelerating urbanization. However, the current reverse logistics system of rural e-business in my country is very imperfect. Obstacles such as reverse logistics without outlets and unreasonable return routes lead to problems such as high logistics costs, difficulty in returning goods, and poor villagers' experience. Therefore, the reverse logistics network of rural e-business is studied. This paper aims at the research of rural e-business logistics and rural reverse logistics, and clarifies the concept of rural e-business reverse logistics by analyzing its concept and characteristics. The problem puts forward construction ideas, and conducts specific research and analysis on the constraint-based clustering algorithm. Using two artificial two-dimensional plane datasets dataset1 and dataset2 to test the clustering effect, dataset1 has 450 data points, divided into 15 groups, each group has 30 data points; dataset2 has 800 data points, divided into 20 groups of 40 data points each. dataset1: Cluster Num=<0, 30>, Clustersize=<0,0.6>; dataset1: Cluster Num=<0, 40>, Clustersize=<0,0.6>. The obtained results are basically reasonable results under the constraints.

**Keywords**

Constrained clustering algorithm; rural e-business; reverse logistics network

## 1. INTRODUCTION

In recent years, the state has been committed to promoting the growth of rural e-business, and has elevated it to the height of promoting rural economic growth, increasing farmers' income and accelerating urbanization[1]. Therefore, the general consumption capacity in rural areas has gradually increased, the enthusiasm of villagers to shop online has been greatly improved, and the demand for rural logistics has continued to expand. Rural e-business logistics is a key link in the growth of rural e-business. At present, my country's rural logistics system is not perfect, and the logistics network is basically not established. The growth of rural e-business is restricted by problems such as lagging in the construction of logistics centers, and it is imminent to build a rural e-business logistics network[2]. At the same time, due to the particularity of e-business platforms, consumers can only check the basic information of products and customer reviews online to decide whether to buy products. The asymmetry of information makes it impossible to guarantee the quality of products, and the poor conditions of some rural roads make it difficult to purchase products. Commodities are damaged or lost in transit, and the frequency of returns and exchanges of rural commodities increases, resulting in reverse logistics. The research on the reverse logistics network of rural e-business is an indispensable link in the construction of rural e-business logistics system[3]. Therefore, this paper conducts research on the rural e-business reverse logistics network based on the constrained clustering algorithm, proposes a clustering algorithm suitable for solving the background problems of the subject, and provides technical support for the automatic division of the guarantee area, which has important

ICCEIC2022@3rd International Conference on Computer Engineering and Intelligent Control
EMAIL: e-mail: 179564855@qq.com (Caihui Guo), e-mail: Khukhmurun@qq.com (T. Khukhmurun), e-mail: 14148545@qq.com (Rongxia Wang), e-mail: 1505782370@qq.com (Lian Dai)

theoretical value and practical value, and emphasizes This paper discusses the importance and necessity of studying the reverse logistics network of rural e-business, and proposes a solution to the existing problems of rural e-business logistics.

## 2. ELEVANT THEORETICAL BASIS

### 2.1. Logistics Network Theory

The logistics network refers to the logistics that are organically combined by the logistics organization network, the logistics infrastructure network and the logistics information network under the conditions of economic integration and information technology, with the pursuit of economies of scale as the core, interconnected by various organizations and facilities Service network system[4]. In this concept, it is clarified that the value core of the logistics network is the pursuit of scale benefits, and the goal of the logistics network is described to make each independent logistics node gather each other, and realize the systemization and socialization step by step. The composition of the logistics network is expounded. It consists of logistics organization network, logistics infrastructure network and logistics information network, as shown in Figure 1.

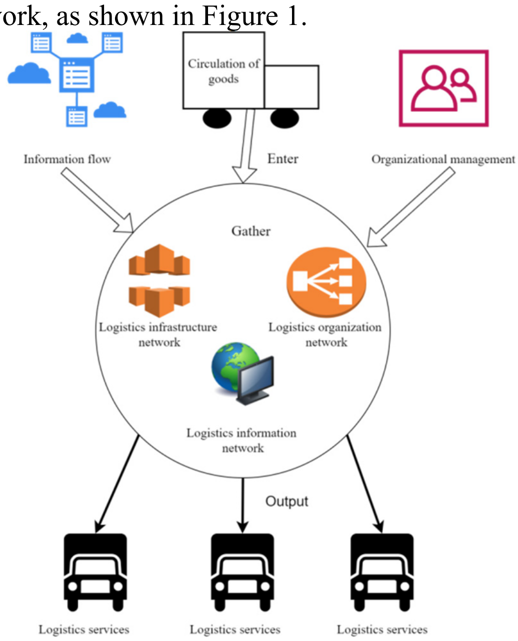


**Figure 1** Logistics network architecture

Its features are as follows:

1) Efficiency of logistics network

The logistics network is an efficient logistics network. Its purpose of operation is to meet the needs of customers with the fastest speed, the fastest speed, and the greatest degree of satisfaction of the required materials in the shortest time and at the lowest cost..

2) The openness of the logistics network

Due to the openness of the logistics network, each node can quickly exchange information with other nodes, and the number of cooperative nodes can be endless, and all nodes can be connected to each other, but the change of a single node will not affect other nodes. run.

3) The leading role of information in the logistics network

The informatization of the logistics network, in addition to the extensive use of mechanized and automated equipment operations, should also use these equipment to collect and manage various information generated in the logistics process, and at the same time conduct in-depth analysis and analysis of the collected logistics information. Mining to make full use of the existing information to guide and integrate the operation of the entire logistics network[5]. In the entire logistics activity, the information flow has the function of guidance and integration; in the entire logistics network, each node responds to the query of the upstream node, sends the service request to the downstream node, and provides logistics services according to the needs and feedback of other nodes. Planning; in material transportation, information flow plays the functions of pre-selection of circulation routes, real-time monitoring and feedback, etc., to guide and integrate the entire logistics process; through virtual scenes and paths, the operation process is greatly simplified, and logistics errors and mistakes are reduced, so that all links of logistics are closely connected, and the interval of each link is greatly shortened.

4) Scale advantage of logistics network

The logistics network pursues economies of scale. It does not take a single node as the core, but organizes scattered nodes into an organic whole, distributes functions to multiple nodes, and forms a networked network through the interconnection of nodes. In the field of logistics, when a node is scattered and management is centralized, its scale advantage will be highlighted. The coordinated operation of large logistics nodes can effectively improve the operation efficiency of the entire logistics system, reduce the overall operating costs, and reduce the dependence on a single node. When a node fails, other nodes can be quickly replaced, thus greatly improving the system's resistance. risk capability.

## 2.2.The concept of rural logistics

The term rural logistics is not common in others countries. It is a concept proposed by many scholars relative to urban logistics after logistics was introduced into China from abroad[6]. At present, there is no unified definition for rural logistics. Some scholars believe that rural logistics refers to the logistics that serve agricultural production and rural residents' life, which serves the needs of agricultural production and rural residents' life[7].

After that, the rural logistics system will be defined, and the rural logistics system will be regarded as a system composed of economic organizations that provide logistics support and services for rural production, life and other economic activities[8]. The rural logistics system is not only related to the characteristics of agricultural production and rural consumption, but also to the current growth of China. In addition, due to the particularity of rural logistics objects, concepts such as rural logistics, agricultural logistics, and agricultural product logistics are easily confused or even equivalent. It is necessary to introduce the concepts of agricultural logistics and agricultural product logistics to distinguish rural logistics, as shown in Table 1.

**Table 1** Comparison of related concepts

| Compare items | Rural logistics | Agricultural logistics | Agricultural product logistics |
|---|---|---|---|
| Agricultural production materials sales logistics (upstream) | Does not include | Include | Does not include |
| Agricultural production materials sales logistics (end) | Include | Include | Does not include |
| Agricultural production logistics | Include | Include | Does not include |
| Agricultural product sales logistics (upstream) | Include | Include | Include |
| Agricultural product sales logistics (downstream) | Does not include | Include | Include |
| Rural consumer goods sales logistics (upstream) | Does not include | Does not include | Does not include |
| Rural consumer goods sales logistics (end) | Include | Does not include | Does not include |

| | | | |
|---|---|---|---|
| Rural Durable Consumer Goods Sales Logistics | Include | Does not include | Does not include |
| Rural infrastructure related logistics | Include | Does not include | Does not include |
| Agriculture-related township and village enterprise logistics | Include | Include | Part includes |
| Non- agricultural township and village enterprise logistics | Include | Does not include | Does not include |

Rural logistics is a concept of regional logistics, which corresponds to urban logistics. As mentioned above, Wang Xinli's definition of rural logistics as centered on agricultural production is rather biased. After all, there are both agricultural production activities and rural consumption activities in rural areas. The new socialist countryside should not only develop agricultural production, but also involve all aspects of rural life.

## 2.3. Rural E-commerce Reverse Logistics

Generally speaking, rural e-business has two broad and narrow concepts. The narrow concept refers to a series of online transactions and management activities carried out around the production and sales of rural agricultural products through network platforms. The broad concept is that as the physical terminal of the rural e-business platform, it is directly rooted in the rural areas to serve the agriculture, rural areas and farmers, realize the implementation of the agriculture, rural areas and farmers, so that farmers become the biggest beneficiaries of the platform, and manage agricultural products in an all-round way based on information technology and network systems. The whole process from origin to consumer[9].

Rural logistics is a reduced concept of regional logistics. Compared with urban logistics, it includes not only the logistics and distribution of top-down consumer goods, but also the outbound sales of agricultural and sideline products produced in rural areas. Its purpose is to serve the production and life of the vast number of rural residents. The entire logistics and distribution process includes a series of narrow logistics activities such as accumulation of goods, space storage, sorting of goods, distribution of goods, and delivery of goods. Commercial activities for the purpose of door-to-door delivery, such as delivery and inspection, are a special and comprehensive logistics distribution link based on the combination of business flow and logistics[10].

This paper defines rural e-business logistics as: physical transportation and distribution activities that occur in rural areas through e-business transactions, sellers and buyers place orders through online shopping platforms, and express delivery, logistics, and bus delivery methods., including agricultural products, agricultural and sideline products, agricultural materials and agricultural tools and daily consumer goods, all kinds of commodities are transported, transited, distributed between the county and township, and finally reach the customer's hands, completing the process of transaction within the rural county.

### 2.3.1. Characteristics of e-business reverse logistics

a) The operation mode is converted from the traditional business flow to the information flow. In the environment of e-business, information flow is the core of reverse logistics, and the real-time monitoring of reverse logistics can be effectively realized by using the transmission function of network information.

b) The source and time of the returned goods are highly uncertain. The demand for reverse logistics is usually a random event, which is passive, difficult to plan and predict for e-business companies, and its source, quantity and frequency are highly uncertain.

c) The return rate is high. Compared with traditional shopping methods, consumers will need to return and exchange goods because they cannot see the real thing, are impulsive, and have personal preferences. At the same time, some merchants publish false information that does not match the actual product, which leads to consumers returning and exchanging goods.

d) Diversification of return channels. In the process of e-business transactions, there are various logistics channels for returns and exchanges. Consumers and merchants can carry out return operations online to improve the efficiency of reverse logistics,as shown in Figure 2.

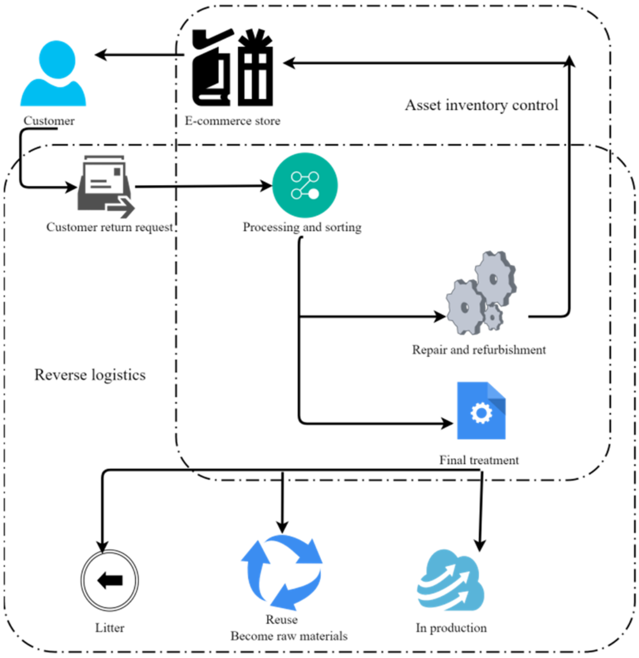


**Figure 2** E-commerce return process

## 2.3.2. Characteristics of rural reverse logistics

a) Dispersion. The geographical distribution of villages and towns is wide and scattered, so the demand for reverse logistics is scattered.

b) Particularity. Most of the rural areas are located in relatively remote locations, with inconvenient transportation, high costs for returning products, waste recycling, transportation, and warehousing.

c) Confusion. Due to the characteristics of the countryside itself, the products returned and exchanged from the countryside are various and different, mainly concentrated in agricultural materials, production and living materials, etc., which need to be classified and inspected.

d) Seasonal. The commodity objects of rural reverse logistics are mostly the purchase of bulky commodities before the festival and the purchase of production materials in the busy season, and different types of agricultural products are closely related to the harvest season.

## 2.4. Clustering Algorithms

Clustering analysis is an important research direction in data mining, and has practical application value in many fields. In recent years, it has been extensively and deeply studied by scholars[11]. The application fields involved in clustering are: biology, statistics, pattern recognition, information retrieval, machine learning, data mining, psychology and other social sciences. Clustering is an important human activity, and people are good at dividing objects into meaningful clusters and assigning specific objects into those clusters.

[Definition 2.1] A cluster is a collection of physical or abstract objects, that is, a grouping in the clustering result, and each grouping is called a cluster.

[Definition 2.2] According to the similarity between objects, the process of $S$ dividing the object set into $\{S_1, S_2, \dots S_k\}$ clusters is called cluster analysis, wherein $S = \{s_i | 0 \le i \le n\}, k \le n$ the division satisfies the following formulas (1), (2), and (3):

$$S_i \ne \phi, S_i \subseteq S(i = 1,2,..., k) \tag{1}$$

$$\bigcup_{i=1}^{k} S_i = S \tag{2}$$

$$S_i \cap S_j = \phi, i, j = 1,2,..., k; i \ne j \tag{3}$$

The description of objects, processes, systems, and relationships in the objective world requires data, which includes various forms, such as text, graphic images, video, and sound. Data is represented by symbols that map entities of interest to some measure, which associates each attribute of the entity with a variable. Assuming that an object has $d$ an attribute, then several $d$ data objects with attributes constitute a $d$ dimensional space. In the $d$ dimensional space, the data objects are called $d$ dimensional data points, and the $d$ dimensional data points $x$ can be expressed as $x = (x_1, x_2, \dots, x_d)$, which $x_i$ represents the $i$ value of the first attribute, which $d$ represents the dimension of the space. A dataset with data objects can be represented as: $n$

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{d1} & x_{d2} & \cdots & x_{dn} \end{bmatrix} \tag{4}$$

[Definition 2.3] Dimension: The dimension of a dataset is the number of attributes of an object in the dataset. Low-dimensional data is qualitatively different from medium- and high-dimensional data.

[Definition 2.4] Sparsity: For some datasets with asymmetric features, most of the attributes of an object have 0 values, and in many cases less than 1% of non-zero items. Sparsity is actually an advantage, since only non-zero values require storage processing.

[Definition 2.5] Resolution: Different data can often be obtained at different resolutions, and the properties of data at different resolutions are also different.

However, clustering with constraints is a research hotspot in recent years, also known as semi-supervised clustering analysis in machine learning. Constraints and domain knowledge can reduce the "blindness" in the clustering process, so the constraint-based clustering method studies how to use a small amount of domain knowledge to improve the performance of clustering algorithms and the availability of clustering results. Domain knowledge can be: the number of objects in a cluster is within a certain range, and some objects must belong to the same cluster or different clusters. The processing methods of domain knowledge can be divided into distance-based methods and constraint-based methods according to the different ways of using them. The distance-based method is to modify the similarity measurement function through domain knowledge, and then use the traditional method for clustering; while the constraint-based method is to integrate the constraints into the process of the algorithm, and modify the process of the traditional clustering algorithm to Restricts the clustering search process so that it can handle constrained clustering problems.

One of the difficulties in clustering with constraints is how to model the constraints. Several different types of constraints have been proposed in the existing literature. Constraints can be divided into the following four categories according to the scope of the object to which the constraints are imposed:

1) Constraints of clustering objects. This constraint restricts the objects participating in the clustering, and selects the objects that meet the conditions for cluster analysis. For example, in a certain clustering, only customers whose consumption amount exceeds 1,000 yuan are clustered. For this constraint, it can be advanced to the preprocessing stage, so that the constrained clustering problem can be transformed into an unconstrained clustering problem.

2) Obstacle constraints. For example, there are rivers, railways, bridges, lakes, mountains, etc. in the city, and these constraints must be considered when clustering objects in the city according to the spatial distance. Such constraints can be de-constrained by redefining the distance metric function, which can then be solved by traditional clustering methods.

3) Parameter constraints. Some constraints can appear in the clustering algorithm in the form of parameters, such as the number of clusters constraints, such constraints are proposed by the user, the

processing is simple, and only exists as a parameter of the algorithm, so such conditions are generally not regarded as constraint.

4) Cluster constraints. The clusters in the clustering results are required to meet certain conditions, such as the deviation of the similarity within the same cluster or the total number of objects contained in the cluster within a certain range. It can also be divided into: a) Rigid constraints, which must be satisfied during clustering; b) Flexible constraints, which should be satisfied to the greatest extent possible during clustering.

The constraints handled by the algorithm in this paper are all rigid constraints

## 3. Using constrained clustering algorithm in rural e-business reverse logistics network

Experimental environment: Intel Pentium(R) 2.8GHz dual-core, memory 1G, Windows XP. An experimental simulation platform is written in Microsoft Visual C#. net environment, which visualizes two-dimensional data points, and outputs the labels after data clustering to a file for other multi-dimensional data sets.

Experimental process: First, two artificial two-dimensional plane datasets dataset1 and dataset2 are used to test the clustering effect. The dissimilarity between data points is measured by Euclidean distance. There are 450 data points in dataset1, which are divided into 15 group, each group has 30 data points; dataset2 has 800 data points, divided into 20 groups, each group has 40 data points. Therefore, the time complexity of the algorithm was tested. Finally, the Iris dataset is used to test the relationship between constraints and clustering accuracy.

The algorithm program can manually create a collection of two-dimensional data points, or it can directly open the data file to read the multi-dimensional data. The two-dimensional data can be visualized to the right panel, while the multi-dimensional data can only save its results to the file. The must Link and cannot Link constraints are implemented by entering data pairs $< v_i, v_j >$, and then clicking on the corresponding data points can create these two types of constraints. Other constraints are similar to the control parameter input from the text box. The attribute constraints and the cluster size constraints are actually equivalent. They are all judged by adding certain conditions when merging subtrees to see if the constraints are violated. Therefore, the simulation experiments of the algorithm do not consider the attribute constraints.

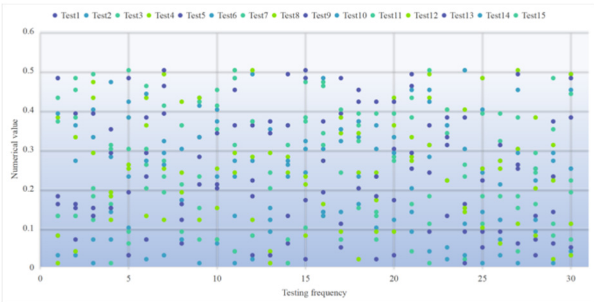


**Figure 3** Distribution of dataset dataset1

On the clustering results of dataset1, the original program displays the class label of the data point next to the point, and distinguishes different clusters by color, because the data points are too dense and the effect is not obvious after black and white printing, so this article will cluster The results are exported to EXCEL, and drawn through EXCEL (as shown in Figure 3), different series represent different clusters and are distinguished by color. Cluster Num=<0, 30>, Clustersize=<0,0.6>, it can be seen from the results in the figure that on the basis of satisfying all constraints, the data set can obtain clusters of arbitrary shape and multi-density, and need to input control parameters only one. The obtained results are basically reasonable results under the constraints.

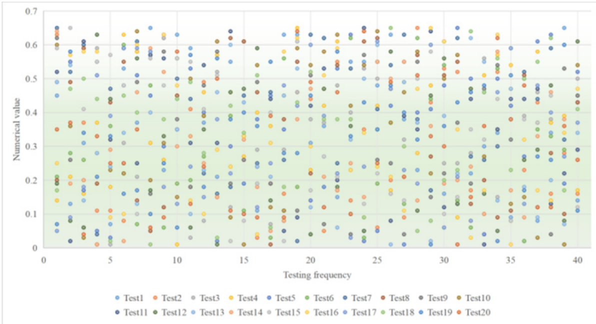


**Figure 4** Distribution of dataset dataset 2

The clustering results of dataset 2 are also exported to EXCEL according to the clustering result mode of dataset 1, and drawn through EXCEL (as shown in Figure 4). Different series represent different clusters and are distinguished by color. Cluster Num=<0, 40>, Clustersize=<0,0.6>, it can be seen from the results in the figure that the dataset also satisfies all constraints, and can also obtain clusters of arbitrary shape and multi-density, and needs There is only one input control parameter. The obtained results are basically reasonable results under the constraints.

## 4. CONCLUSION

In recent years, the state has been committed to promoting the growth of rural e-business, and has raised it to the level of promoting rural economic growth, increasing farmers' income and accelerating urbanization. The construction of rural e-business logistics system has achieved initial results. In rural e-business transactions, when villagers buy daily consumer goods and sell agricultural products through e-business platforms, the phenomenon of "return and exchange" will inevitably occur, resulting in reverse logistics. However, the current reverse logistics system of rural e-business in my country is very imperfect. Obstacles such as reverse logistics without outlets and unreasonable return routes lead to high logistics costs, difficulty in returning goods, and poor villagers' experience, which restricts the growth of rural e-business. Firstly, on the basis of expounding the related concepts and characteristics of rural e-business reverse logistics, this paper studies the current situation of rural e-business and logistics, and conducts research through constrained cluster analysis, which provides important theoretical support for this research. And put forward some countermeasures:

1)Scientifically plan the layout of rural e-business logistics network;
2)Actively implement the sharing of rural e-business logistics facilities;
3)Improve the operation standards of agricultural products e-business enterprises;
4)Further improve the mandatory standards for cold chain logistics of agricultural products;
5)Actively cultivate the main body of rural e-business logistics operation;
6)Speed up the innovation of rural e-business logistics operation mode.

## 5.ACKNOWLEDGMENTS：

【Fund project1】This article is based on 2018 Guangzhou Nanyang Polytechnic Polytechnic "Innovation Strong School Project" Project (NY-2022CQ-JGW004)

【Fund project2】This article is based on the key research platform project of education department of Guangdong Province(2017GGXJK103)

## 6.REFERENCES:

[1] Guo L. Accelerate the Preliminary Study on Informationization of Urbanization Development. Urban Studies, vol.23, no.14, pp.125-129,2010.

[2] Liu S. On Promoting the Construction of Logistics Network System of Rural E-commerce. Journal of Heze University, vol.18, no.9, pp.390-393,2019.

[3] Fuyan, Yang, Xiaodong. Research on Development Status and Problems of Rural E-Commerce Logistics Management & Technology of Sme, vol.26, no.16, pp.656-658,2018.
[4] Qin L, Hu J. A Model of Price Determining in Intelligent System of Logistics Cloud-Network. Fifth International Conference on Measuring Technology & Mechatronics Automation. IEEE Computer Society, vol.9, no.4, pp.1789-1789,2013.
[5] Zhu C Z, Jun-Suo Q U. Analysis of China logistics informatization growth[J]. Journal of Xi'an University of Posts and Telecommunications, vol.8, no.4, pp.189-193,2010.
[6] Hai-Jun H E. China's logistics growth in rural areas and urban logistics Comparison and Implications. Logistics Engineering and Management, vol.48, no.39, pp.4895-4891,2010.
[7] HU, Cui-hong*, Southwest. Rural Logistics System Based on Rural Informatization. Asian Agricultural Research, vol.5, no.2, pp.4-4,2010.
[8] Chen D L. Research on Dynamics Mechanism of Constructing Rural Logistics System. China Business and Market, vol.15, no.7, pp.179-182,2011.
[9] Xiong H. New Exploration of System Construction about Agricultural Products Traceability Origins From the First Layer to the Second Layer. Future and Development, vol.21, no.19, pp.387-394,2010.
[10] Li-Xiao L I. Study on the Influencing Factors on China's Rural Logistics and Its Development Countermeasures. Prices Monthly, vol.17, no.14, pp.150-154,2015.
[11] Ben D, Pitskel N B, Pelphrey K A. Three Systems of Insular Functional Connectivity Identified with Cluster Analysis. Cerebral Cortex, vol.21, no.7, pp.1498-1506,2010.