International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Multi-Factor Model Base on the Minimal Spanning Tree

Xiangkun Zheng

School of Sciences, Hebei University of Technology
School of Sciences, Hebei University of Technology, Beichen district,Tianjin 300401, China.

Abstract: ***In this article, through Kruskal and Prim algorithm to solve minimum spanning tree, the application of group technology. by establishing the model, Solve the problem of model, the final result applied to of many factors to pick stocks in the model, we can find their contact more clearly.***

Keywords: kruskal, prim, the minimal spanning tree, multi-factor model

1. Introduction

For reasonable pricing of stocks and other financial assets is the main content of financial theory research. Research of shares and other securities assets pricing to explain and predict the price of the subject matter of the future is of great significance. In stock, for example, at present in China's stock market, analysis of market price movements of prediction methods are divided into two schools:

One is technical analysis of genre, technical analysis focuses on the dynamics of the market, the history of stock price movements. Technical analysis will be a lot of energy into historical market research Gui, based on the history could repeat itself this hypothesis, through the chart analysis or index analysis to find the price on the basis of the laws and historical trends to predict the future of the stock price movements. Technical analysis of the advantage is that can tell investors should be in what time to buy stocks or sell stocks. Technical analysis, however, depends on the history repeats itself, because the stock market is changing in reality, the effectiveness of technical analysis is often challenged, it is also the limitations of technical analysis.

Another genre is fundamental analysis. This genre, mainly from the perspective of economic theory, using the more perfect macro analysis, industry analysis and corporate earnings analysis, extract decided to valuations of useful information to analyze the stock market. Fundamental to pick stocks based on growth of the stock, and whether the value of the stock is undervalued, and so on and so forth. In addition, the fundamentals of the stock market also need to consider some dynamic information, including the policy information and industry information for analysis. Fundamental analysis can help investors choose the better, but for the timing of the stock, but there was no clear answer.

Quantitative investment can be combined the advantages of both. Through the quantitative concept originally belong to the category of qualitative analysis is the fundamental analysis of quantitative, at the same time combined with technical analysis itself has the characteristics of quantitative analysis, quantitative investment can combine the two together, can simultaneously satisfy the investors demand for stock selection and timing. And more factor model is the combination of the two analysis methods in a typical quantitative investment model. Research on multiple factor actually has been for some time, Fama and three factor model put forward by the French to illustrate the factors of affecting stocks, the current domestic brokers.

Clustering analysis is an important technology in data mining, as people gradually in-depth understanding of clustering analysis and clustering is widely used in production and living, more and more researchers have realized the significance of clustering and clustering in the field of computer science and technology has become a large application research hot spot. Clustering analysis comes from the core idea of birds of a feather flock together, on the basis of the data object attributes will be divided into several classes and objects, makes the object in the same class as much as possible, and not same as possible between the objects. Cluster analysis, the original exploration with little or no prior knowledge, including research and development in.

2. The Problem Minimum Spanning Tree

Problem: in real life there are many similarities between towns set up telephone lines, pipelines, building roads, usually in the early stage of development, due to technical or financial limitations, people always from the perspective of saving material or money, try to design a network to make different towns can be connected directly or indirectly, the total length of the shortest.

Several complementary concepts:
Undirected graph $G(V,E)$ and directed graph $D(V,E)$; Subgraph; (undirected graph) chain with primary chain, circle and primary coil; (directed graph) with primary road, circuits and primary circuit; Connected graph (these concepts or refer to the graph modeling method or reference specialized books);

Tree: undirected graph $G(V,E)$ is a connected graph of acyclic, call it a tree;

The nature of the tree:

Tree number is equal to the edge of the vertex number minus "1", namely $|E|=|V|-1$;

Tree between any two vertices connected just have a primary chain;

After removing any an edge in the tree, then get a disconnected graph;

Between any two vertices in the tree to add a new edge, the income of it just have a primary coil.

Spanning tree: if the figure G is the tree of the generated graph H, it is called H is the spanning tree of G.

Generally speaking, a spanning tree is not only a figure.

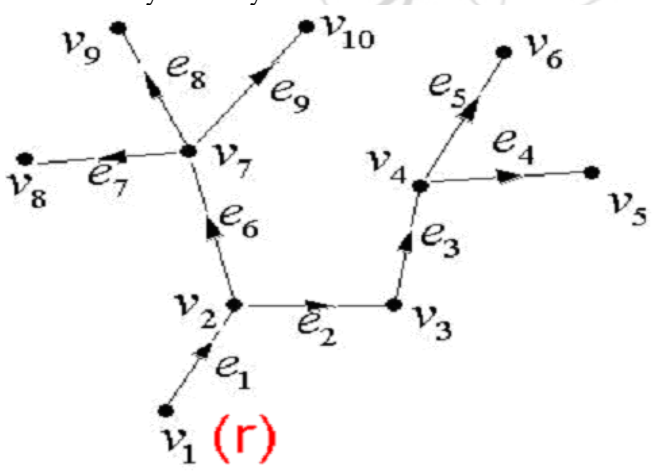
3. Minimum Spanning Tree

Tree root: directed graph $D(V,E)$, if $\exists r \in V$, $\forall v \in V$, there is only one from the vertex root r to the road v, it is said to the directed graph $D(V,E)$ is a tree, vertex to the root.

The nature of the root tree $D(V,E)$:

For $\forall v \in V$,There is only one from the root r to the vertex V way;

Root tree $D(V,E)$ relative to its foundation drawing (regardless of the direction of the edge, $D(V,E)$ the corresponding undirected graph) tree $G(V,E)$, it determined by root only.



In an empowerment connected graph $G(V,E;w(\cdot))$ to find a spanning tree T (T edge set for E_1), makes the combined income and the weight of the each side of the tree $w(T)=\sum w(e)$ (also known as the right) is minimal. Called G is the right of the minimum spanning tree T^* in the minimum spanning tree G.

Theorem spanning tree must be connected graph.

4. Minimum Spanning Tree Algorithm

A simple connected graph as long as it's not a tree, its spanning tree is not the only, and very much. Generally, complete graph *n* vertices, the different number for spanning tree is n^{n-2}. Thus, for a given empowerment figure of minimum spanning tree, usually can't use the exhaustive method. The vertex complete graph, for example, a spanning tree have 30 vertices, have 30^{28}spanning tree, which uses the most modern computer, in our lifetime is not exhaustive. So, exhaustive method algorithm for minimum spanning tree is invalid must seek effective algorithm.

In the effective algorithm for solving the minimum spanning tree, two of the most famous is the Kruskal algorithm and Prim algorithm, the iterative process is to design based on the greedy method.

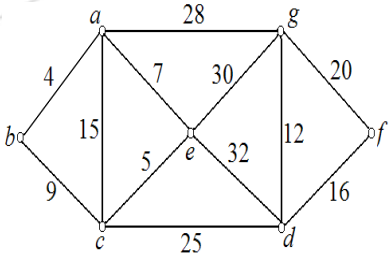
5. Kruskal Minimum Spanning Tree Algorithm

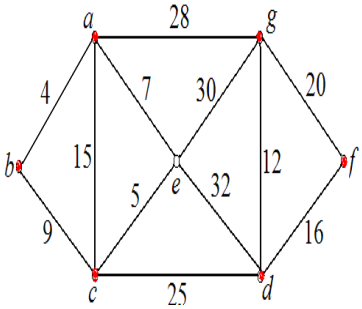
5.1 The description of Kruskal intuitive algorithm

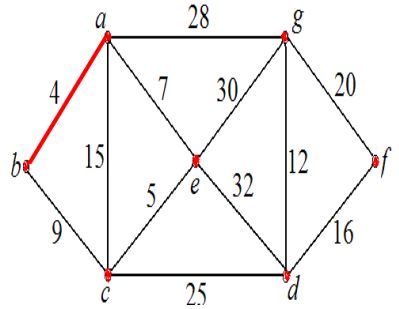
Hypothesis T_0 is that the minimum spanning tree of *G* empowerment diagram, the edges of *G* vertices are painted red, initial the edges are white.

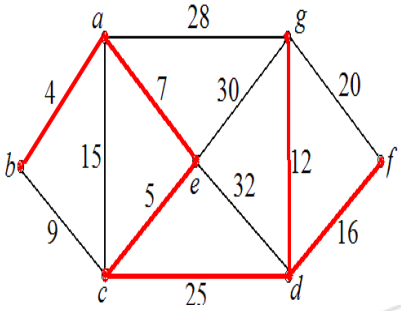
1) all the vertices painted red;
2) choose a right minimum value at the edge of the white edge, make it do not form a circle with red edge, the white edge TuGong;
3) repeat (2) until $n-1$ edge is a red edge, the $n-1$ red edge constitute the set of T_0 minimum spanning tree edges.

For example, for the image below (a) shows the empowerment of the figure, according to the above steps, easy to find out the minimum spanning tree. Among them,(b) (c) is the first step and step 2 respectively, (d) is the final result.









5.2 Kruskal Algorithm

Right $G(V,E)$ to the edge m by increasing the order, then check whether each edge is selected in turn. Set the current to check for the edge e_k, just collection is composed of the selected edge E_1. If $G_2=(V,E_1\cup\{e_k\}$ no circle in the figure, then e_k is chosen by E_1; Otherwise don't choose by e_k, and to continue to check an edge.

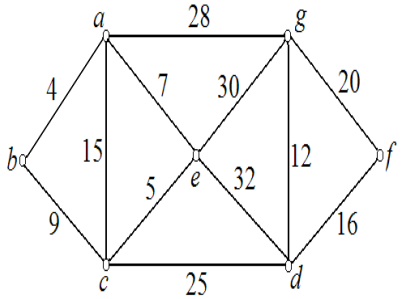
To G the composition of each vertex V based on the E_1 current edge condition defined a label $l(v)$, made with features: if $G(V,E_1)$ and only if there is a connection in the picture the vertices u and v chain, there are $l(u)=l(v)$.

theorem If the edge of the current inspection $e_k=[u,v]$ does not belong to E_1, in the figure $l(u)=l(v)$,then $G_2=(V,E_1\cup\{e_k\})$ must be in the ring; In $l(u)\neq l(v)$, G_2 must have no ring.

5.3 Sample Application

Example,use Kruskal algorithm for below given the empowerment of the minimum spanning tree of the graph

Solution: the edges of the figure in accordance with the weight since the childhood are:

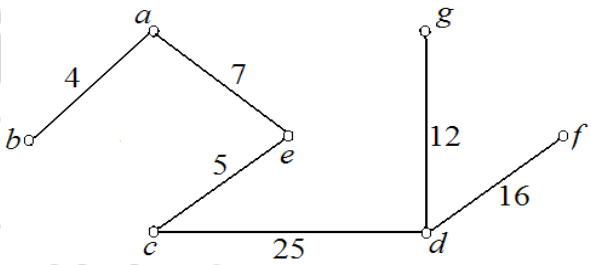


边	(a,b)	(c,e)	(a,e)	(b,c)	(d,g)	(a,c)
权	4	5	7	9	12	15
边	(d,f)	(f,g)	(c,d)	(a,g)	(e,g)	(d,e)
权	16	20	25	28	30	32

Operating Kruskal algorithm, iterative step 9 complete minimum spanning tree search. Operation process of the steps listed in the table below.

步骤	选出边e	w(e)	操作	VS	T_0	$C(T_0)$
1	(a, b)	4	加到T_0中	{{a, b}, {c}, {d}, {e}, {f}, {g}}	{(a, b)}	4
2	(c, e)	5	加到T_0中	{{a, b}, {c, e}, {d}, {f}, {g}}	{(a, b), (c, e)}	9
3	(a, e)	7	加到T_0中	{{a, b, c, e}, {d}, {f}, {g}}	{(a, b), (c, e), (a, e)}	16
4	(b, c)	9	删除	{{a, b, c, e}, {d}, {f}, {g}}	{(a, b), (c, e), (a, e)}	16
5	(d, g)	12	加到T_0中	{{a, b, c, e}, {d, g}, {f}}	{(a, b), (c, e), (a, e), (d, g)}	28
6	(a, c)	15	删除	{{a, b, c, e}, {d, g}, {f}}	{(a, b), (c, e), (a, e), (d, g)}	28
7	(d, f)	16	加到T_0中	{{a, b, c, e}, {d, g, f}}	{(a, b), (c, e), (a, e), (d, g), (d, f)}	44
8	(f, g)	20	删除	{{a, b, c, e}, {d, g, f}}	{(a, b), (c, e), (a, e), (d, g), (d, f), (c, d)}	44
9	(c, d)	25	加到T_0中	{{a, b, c, e, d, g, f}}	{(a, b), (c, e), (a, e), (d, g), (d, f), (c, d)}	69

The results shown in figure



6. To Strive for the Minimum Spanning Tree Prim Algorithm

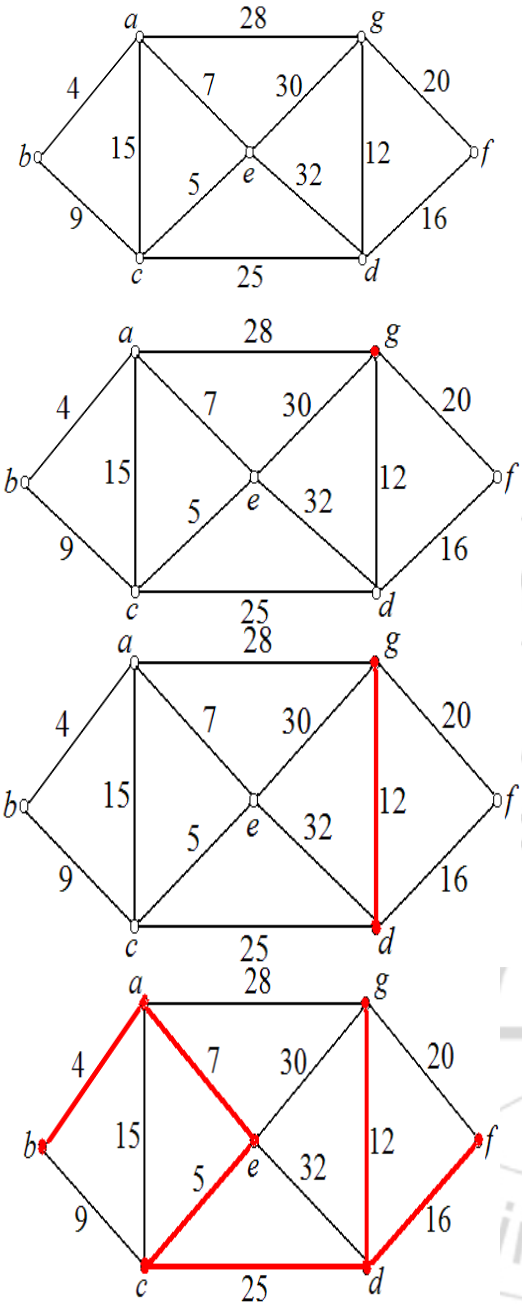
6.1 Intuitive description of the algorithm

Assumption T_0 is empowerment minimum spanning tree of the graph. Choose a vertex TuGong, the rest of the vertices is white; On an endpoint for the red, the other one endpoint for the edge of white, look for a right of the smallest side.

TuGong, the white, on the edge of the endpoint also painted red. So, every time an edge and a vertex painted red, until all the vertices into red. Finally red edge constitute the set of minimum spanning tree edges T_0

For example, for the image below (a) shows the empowerment of the figure, according to the above description, easy to find out the minimum spanning tree.

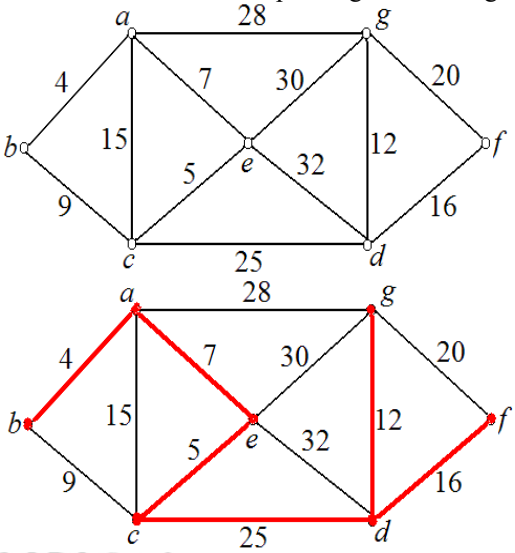
Below, (b)、(c) is the first step and step 2 respectively, (d) is the final result.



6.2 Basic Idea of the Algorithm

Set T_0 and $C(T_0)$ the edges of the figure G of the minimum spanning tree respectively set and its weights, the initial state are empty, at the end of the algorithm T_0 contains all the side of the minimum spanning tree, $C(T_0)$ is said the weight of the minimum spanning tree. First specify a vertex for initial access point v_0, remember v_0 to do it, will be added to the set of V', and then find the jumper in the collection "by point" and "failed" collection right between the smallest side as "by side" to join in T_0, and $V - V'$ will in turn to the V' endpoint. Repeat the process until $V' = V$ so far.

Patients using the Prim algorithm for below the empowerment of the minimum spanning tree of the graph.



Solution: to keep things simple, the operation Prim algorithm steps listed in the table below.

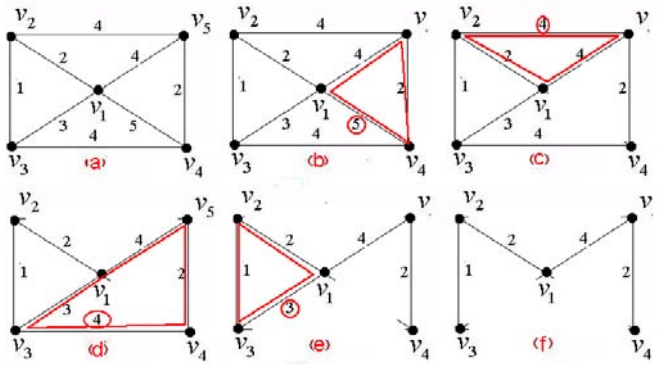
步骤	u	L(b)	L(c)	L(d)	L(e)	L(f)	L(g)	e	V'	T_0	$C(T_0)$
1	a	4	15	∞	7	∞	28		{a}	∅	0
2	b	–	9	∞	7	∞	28	(a,b)	{a, b}	{(a, b)}	4
3	e	–	5	32	–	∞	28	(a,e)	{a, b, e}	{(a, b), (a, e)}	11
4	c	–	–	25	–	∞	28	(c,e)	{a, b, e, c}	{(a, b), (a, e), (c, e)}	16
5	d	–	–	–	–	16	12	(c,d)	{a, b, e, c, d}	{(a, b), (a, e), (c, e), (c, d)}	41
6	g	–	–	–	–	16	–	(d,g)	{a, b, e, c, d, g}	{(a, b), (a, e), (c, e), (c, d), (d, g)}	53
7	f	–	–	–	–	–	–	(d, f)	{a, b, e, c, d, g, f}	{(a, b), (a, e), (c, e), (c, d), (d, g), (d, f)}	69

7. Broken Ring Method

The ideas of the algorithm is the right of all edges of the figure in accordance with the edge size from big to small order, on the basis of the original image, once upon a time to inspect the edge, after every inspection, verify whether there is including the edge of the primary coil. If yes, will be deleted from the side concentration, otherwise remain. Until the rest of the children the graph is a tree (you can take the rest of the number of edges as the test conditions), or an edge under investigation.

According to this thought structure to solve the minimum spanning tree algorithm is referred to as "broken ring method" (the *Kruskal* algorithm to avoid circle method "), you can echo *Kruskal* algorithm gives accurate portrayal of "broken ring method".

The following picture is the use of broken circle method to solve the concrete implementation of the above example:



Group technology: group technology is a method of design and manufacturing system, it put the production parts of the machine group, the classification of parts to production accordingly, about a situation in which across a set of processing parts as far as possible, ideally each parts are made within the group.

机器	1	2	3	4	5	6	7	8	9
加工的零件	2，3，7，8，9，12，13	2，7，8，11，12	1，6	3，5，10	3，7，8，9，12，13	5	4，10	4，10	6

Assuming that there are 13 kinds of parts, in 9 machine processing. In the part number of each machine processing is given in the table below.

Will this 3 to 9 machine group, make the parts processing situation as far as possible across the group.

Solution: (1)modeling: M_i means collection of parts to processed by machines. For any two machines i, j, definition i and j phase thin place:

$$w(i,j)=\frac{\left|M_i \oplus M_j\right|}{\left|M_i \cup M_j\right|}$$

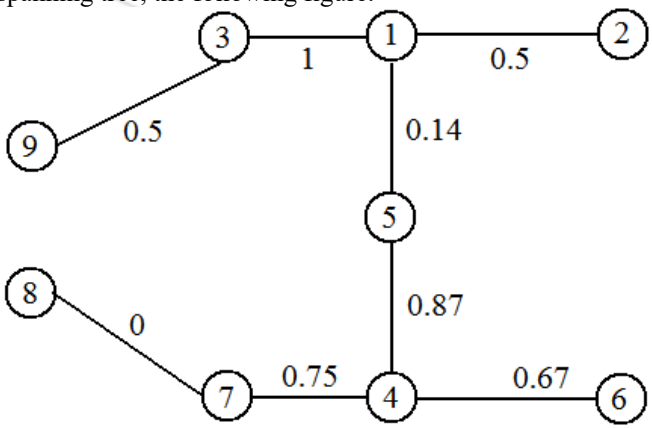
The symmetry $\oplus$ is poor,said molecule is in the machine i but not on the machine j processing, or number of parts in the machine j but not on the machine i processing. The denominator for the processing of parts on the machine i or the machine j number. Obviously $0 \le w \le 1$, $w(i,j)=0$ is said machine i and machine j parts are exactly the same; $w(i,j)=0$ is said Machines i and machine j parts are not the same. w expression of two different dissimilarity degree of machine parts.

In the machine for the vertices, a complete graph, each edge (i,j) is given $w(i,j)$. Minimum spanning tree of the graph is composed of those thin the smallest edge con, or as to get rid of the phase "thin is relatively large side after the rest of the connected graph. If you want to put the machine into a group k, you continue to delete the biggest side right of minimum spanning tree $k-1$. So get a separate subtrees k, the vertices of each tree into the machine group.

(2)solution method for model the data given in the table in front of, in accordance with the above modeling method to construct weighted graph, edge weight matrix shown in the following table.

边	1	1	1	1	1	1	1	1	2	2	2	2
	2	3	4	5	6	7	8	9	3	4	5	6
边权	0.5	1	0.89	0.14	1	1	1	1	1	1	0.62	1
边	2	2	2	3	3	3	3	3	3	4	4	4
	7	8	9	4	5	6	7	8	9	5	6	7
边权	1	1	1	1	1	1	1	1	0.5	0.87	0.67	0.75
边	4	4	5	5	5	5	6	6	6	7	7	8
	8	9	6	7	8	9	7	8	9	8	9	9
边权	0.75	1	1	1	1	1	1	1	1	0	1	1

With the *Kruskal* algorithm can calculate the minimum spanning tree, the following figure.



The two edges of the most power in the minimum spanning tree, get three separate trees, their vertices are: $\{3,9\},\{1,2,5\},\{4,6,7,8\}$ this is the grouping of the machine.

Multiple factor model is the most widely used a stock selection model, the basic principle is to use a series of factors as stock selection criteria, meet the stock of these factors are buying, does not meet the sell.

The basic concept
A simple example: if there is a group of people take part in the marathon, want to know who would be running to the grade point average, it only need to do before running a physical test. The health index of the athletes, more likely to get beyond the average scores. The principle of multiple factor model is similar, as long as we find that yield the most relevant to the enterprise factor.

Various factor model is the core of the difference between the first it is more on the factors of selecting, the second is on how to use the multi-factor comprehensive get a final judgment.

We can put the nine kinds of machines as nine stocks respectively, which kind of parts processing, respectively, as a factor, can according to the *Kruskal* algorithm to find the minimum spanning tree, then the most power in the

minimum spanning tree on either side of the take out, got the separation of the three trees, their vertices are $\{3,9\},\{1,2,5\},\{4,6,7,8\}$: we see the classes of shares 3 and the shares of 9 the factor is similar to that of the first class, the second class and the type of stock 5 factor is similar,that the fourth class, the sixth class, the seventh class and the eighth class of factor is similar.

8. Conclusion

Multifactor model is one of the current international mainstream quantitative investment model, is also a hot issue in the field of quantitative investment in China at present. Multifactor model through modeling method for driving the stock market the explanation and analysis on the factors of price change. Multifactor model research Gui will also be the brokerage and investment fund operation has certain guiding significance. Popular at home and abroad, many quantitative investment model is to build a multiple factor model based on the framework, so research Gui factor model more efficient modeling method is quantitative investment is an important problem in the trade. This paper, based on the method of solving the minimum spanning tree, to solve the factor of the links between different stocks, a stock and make contact between each factor, the difference is more clear, has reached the desired effect.

References

[1] Ding Peng quantitative investment - Strategies and technology [M] Beijing, Electronic Industry Press, 2012.4.
[2] Cai Jianlin. China A-share market, stock selection model [D]. Shanghai Jiaotong University, 2009.
[3] An Empirical Study of Liu Yi factor model in China stock market [D] Shanghai: Fudan University, 2012.
[4] Improvement the minimum spanning tree clustering Xie Zhiqiang, in bright, Yang Jing cube algorithm [J] Harbin Engineering University, 2008,29 (8): 851-857.
[5] Zhou Shibing clustering analysis method for determining the optimum number of clusters Research and Application of [D]. Jiangnan University, 2011.
[6] J. Geralds, "Sega Ends Production of Dreamcast," vnunet.com, para. 2, Jan. 31, 2001. [Online]. Available: http://nl1.vnunet.com/news/1116995. [Accessed: Sept. 12, 2004]. (General Internet site)
[7] John Lintner (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets [J]， Review of Economics and Statistics.