Acta Technica 62 (2017), No. 5B, 531–538

Research on the correlation of cloud computing enterprises' technology innovation layout

Wan Gangyong [1,2], Liu Xianming [1]

Abstract. In order to rapidly and accurately develop a more suitable cloud computing technology innovation strategy for the future, the enterprises need to understand the technology innovation layout of the same industry, which is called the mutant ever victorious. This paper chose the enterprises with over seven invention patents of cloud computing from United States Patent and Trademark Office's (USPTO) patent database as the research object, calculated the similarity of enterprises' technological resource allocation using weighted bipartite graph projection algorithm. Then Pearson correlation analysis was adopted to calculate the correlation coefficient of enterprises' technology innovation layout from the perspective of cloud computing industry. The results show that there are three kinds of technology innovation correlation in this field. With the three kinds of correlation, potential competitors and future partners are excavated. The results obtained in this study are expected to provide valuable information and reference for enterprises' technology innovation layout in the future.

Key words. cloud computing; technology innovation layout; weighted bipartite graph; Pearson correlation.

1. Introduction

Since 2006 when the Google's CEO Eric first proposed the concept of "cloud computing" in the search engine conference, the information technology companies have researched and applied the cloud computing. Cloud computing, as a new technology revolution, has pushed human society forward to a new phase, deeply affected the industry development direction, information interaction mode and security management mode. As a new business model, the cloud computing certainly has a profound impact on the national industrial strategy and the core competitiveness of countries and enterprises. Therefore, how to fully grasp the development trend of cloud com-

[1]Workshop 1 - Information and Communication Branch, State Grid Jiangxi Electric Power Company, 330077, China

[2]Corresponding author: Wan Gangyong; e-mail: 13870806282@163.com

puting and understand the technology innovation layout of enterprises in the cloud computing industry have become one of the problems to be solved.

Therefore, this study explores the correlation of technology innovation layout in cloud computing industry, by analyzing the cloud computing patent data of USPTO, to help enterprises identify competitors, excavate potential partners and provide the basis for promoting the cloud computing technology innovation and hence making more effective technology innovation strategy.

2. Literature review

Technology proximity is the important basis for enterprises to search the potential partners and identify the competitors [1]. Katrin studied the merger cases in Germany in the 90s, and found that enterprises with high technology proximity were often merged [2]. Sears and Hoetker also proposed that the merger effect was positively related to the degree of enterprise technology proximity [3].

Patent is widely used to measure the level of enterprises' technology innovation, identify potential technology and provide basis for technological innovation decision [4]. Using the patent data could analyze the technological innovation trend. At present, the methods, which measure the technology proximity with patent information, are mainly divided into two categories: one is based on the patent citation; the other is based on IPC [5]. Chang [6] explored the linkage between enterprises, and found the path of technology diffusion through analyzed the patent citation. Yoon and Kim pointed out that the new patents were not cited frequently in most cases, so the new technology could not attract wide attention [7]. Hence while the patent citation analysis method could reveal the degree of technology proximity, it often lagged behind. In addition, Lo also found that what the analysis results of patent citation reflected was the correlation between core technologies, so using this method was unable to describe the overall technological innovation strategic layout [8]. The patent citation analysis method has some defects, so many scholars use IPC code to measure the technology proximity. Angue et al divided the enterprise patents into the basic technology and special technology, and used the multilevel IPC to analyze the similarity of enterprises' technology innovation layout in these two technologies to identify the potential partners [9]. Hong Yong et al examined the degree of technology proximity by summarizing different enterprises' patent number in the same IPC code [10].

To sum up, using technology proximity to identify the difference of technology innovation strategy layout and excavate the relationship between enterprises has obtained certain achievements, but it still exist many defects. Previous studies only consider the absolute number of patents held by enterprise, but ignore the enterprises' total technological resources and technology innovation tendency.

3. Materials and methods

3.1. Data acquisition and pre-processing

With the keywords " cloud computing" OR "cloud platform" OR "cloud technology" OR "cloud storage" OR "cloud service" OR "cloud data" OR "cloud terminal" OR "cloud system" OR "cloud security" OR "Platform as a Service" OR Paas OR "Software as a Service" OR Saas OR "Infrastructure as a Service" OR Iaas OR "private cloud" OR "community cloud" OR "public cloud" OR "hybrid cloud" [11], a total of 6996 innovation patents in 2000-2016 were retrieved. Then, the raw data was pre-processed. After eliminating the duplicate data and those that did not belong to cloud computing, finally, we got a total of 2859 patent records, including 113 4-digit IPC codes, and a total of 698 patentees (including enterprises and individuals). In order to avoid an interference caused by excessive number of enterprises, this paper selects 66 enterprises with more than 7 patents that are applied every two years. The top 5 enterprises' names are denoted by serial numbers and separately shown in table 1 below.

Table 1 the list of top 5 enterprises

Number	Name
1	International Business Machines Corporation
2	Microsoft Corporation
3	Red Hat, Inc.
4	Blaze Mobile, Inc.
5	Samsung Electronics, Co., Led.

3.2. Methods

Bipartite Graph is often used to describe the complex relationship between two different kinds of things [12]. Though we can know whether there is a relationship between two things with the help of bipartite graph, it cannot accurately reflect the degree of relationship, so the weighted bipartite graph is put forward on the basis of bipartite graph, defined as follows.

Set the weighted bipartite graph $G(X, Y, E, P)$, X and Y are respectively two types of mutually disjoint peaks in graph G. X is a set of top nodes with the number of n,and Y is a set of bottom nodes with the number of $m.E$ is the set of edges that only exists between the nodes on the top and at the bottom. P is the $n \times m$ edge weighted matrix, let e_{ij} as the line from the node i in X to the node j in Y, then p_{ij} is the corresponding edge weight of the line e_{ij}, and

$$p_{ij} \begin{cases} \neq 0 e_{ij} = 1 \\ = 0 e_{ij} = 0 \end{cases} \tag{1}$$

Compared to the relationship between two different kinds of nodes, studying the relationship between the same kinds of nodes has more application value.

The projection process is divided into two steps. First, conduct the compression projection for the nodes in X of the weighted bipartite graph$G(X,Y,E,P)$, and allocate the information to each node in Y according to the edge weight proportion. Then conduct the compression projection for each node in Y, and allocate the information to each node in X according to the edge weight proportion. Finally, obtain a $n \times n$ matrix composed of node information in X, the values of which represent the relationship weight between nodes on the top.

This process is represented by the symbol as follows:

Set the weighted bipartite graph $G(X,Y,E,P)$, the resources held by nodes of X as $f(X)$, and the resources held by nodes of Y as $f(Y)$. First of all, conduct the projection compression for the resource of nodes in X to the nodes in Y. Here let $f'(Y)$ represent the resource that obtained by nodes of Y, and

$$f'(y_j) = \sum_{i=1}^{n} \frac{p_{ij} e_{ij} f(x_i)}{\sum\limits_{j=1}^{m} p_{ij}} \tag{2}$$

Then conduct the projection compression for the resource $f'(Y)$ to the nodes in X. Here let $f'(X)$ represent the resource that obtained by nodes of X, and

$$f'(x_i) = \sum_{j=1}^{m} \frac{p_{ij} e_{ij} f'(y_j)}{\sum\limits_{i=1}^{n} p_{ij}} = \sum_{j=1}^{m} \frac{p_{ij} e_{ij}}{\sum\limits_{i=1}^{n} p_{ij}} \cdot \sum_{i=1}^{n} \frac{p_{ij} e_{ij}}{\sum\limits_{j=1}^{m} p_{ij}} f(x_i) \tag{3}$$

4. Results

4.1. Similarity of cloud computing enterprises' resource allocation

Use the filtered data to construct the weighted bipartite graph between the enterprise and IPC codes. Set the enterprises as the top nodes of this weighted bipartite graph with the nodes number of 66, IPC codes as the bottom nodes with the nodes number of 113. When the enterprise has applies for the patent in a field, the enterprise and the IPC code on behalf of the field is connected, and the patent number applied by enterprises in this field denotes the edge weight.

If the original nodes are seen as the eigenvectors of target node, the corresponding resource allocation similarity coefficient is the eigenvalue of target node. At this point, the Pearson correlation coefficient algorithm is used to calculate the target nodes' correlation coefficient of technology innovation layout in the industry and explore the relationship between enterprises, with the cloud computing industry being seen as a system composed of the target nodes.

Use SPSS software to calculate the Pearson correlation coefficient among 66 target nodes, and a 66×66 Pearson correlation coefficient matrix is obtained. Due to the space limitation, the Pearson correlation coefficient matrix of the enterprises whose total number of patents in top 5 is listed in the table1 only.

Table 1 the Pearson correlation coefficient matrix of the enterprises whose total number of patents in top 5

	1	2	3	4	5
1	1.000	0.358	0.384	-0.230	-0.818
2	0.358	1.000	0.150	-0.220	-0.033
3	0.384	0.150	1.000	-0.196	-0.207
4	-0.230	-0.220	-0.196	1.000	-0.120
5	-0.818	-0.033	-0.207	-0.120	1.000

It can be seen from the above table that there is weakly negative correlation among Blaze Mobile, Inc. and other nine enterprises. According to the definition of Pearson correlation coefficient, it can be understood that the technology innovation layout of Blaze Mobile, Inc. has no correlation with other nine enterprises. Because the correlation coefficient of technology innovation layout between International Business Machines Corporation and Samsung Electronics, Co., Ltd. is -0.818, the technology innovation layouts of these two enterprises are complementary.

4.2. Identify the correlation and relationship of enterprises'technology innovation layout

There are a number of enterprises involved in this paper, and their relationships are complex, so in order to understand the correlation of technology innovation layout of 66 enterprises more intuitively, this paper adopts Pajek software to draw the relationship network among the 66 enterprises. The detail of the network is as shown in figure 1 below. In this network, the node represents the enterprise, and the edge weight between two nodes represents the correlation coefficient of technology innovation layout between these two enterprises. In order to seize the main relationship between enterprises, according to the classification standard of Pearson correlation coefficient, a complex network is built with the edges whose absolute value of weight greater than |0.4|. The full line means positive relationship between two nodes, while the dotted line means negative relationship between two nodes in this complex network, and the node name is as shown in the table 1.

The analysis for the three kinds of correlation of technology innovation layouts in figure 1 is as follows:

First of all, there is mainly positive correlation between nodes. Through analyzing the relationship between enterprises with such characteristics, it can be found that the technology innovation layout is highly overlapping.

Except those enterprises with cooperative relationship (e.g., Hon Hai Precision Industry Co., Ltd. and Hong Fu Jin Precision Industry (ShenZhen) Co., Ltd. have significant positive correlation on their technology innovation layouts due to the co-

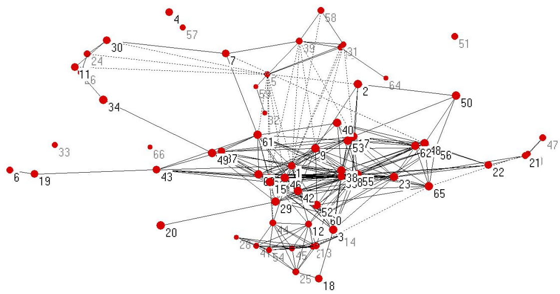

Fig. 1. the correlation network of technology innovation layout

operation), the technology has high similarity, so it is easier to form the competitive relationship between enterprises. In addition, when the correlation of technology innovation layout is higher and the enterprise scale is much different, the smaller enterprise will be merged more easily. The research of Katrin, Sears and Hoetker also support this conclusion in this paper. It obviously can be seen that how to deal the positive correlation of technology innovation layout with counterparts, timely grasp the mainstream technology and reduce the competition has become the key issue for enterprises.

Secondly, there is a little negative correlation between nodes. For example, Samsung Electronics, Co., Led. , Hon Hai Precision Industry Co., Ltd. , Hong Fu Jin Precision Industry (ShenZhen) Co., Ltd. , Harris Corporation and Kabushiki Kaisha Topcon all have significant negative correlation with International Business Machines Corporation. It indicates that the technology innovation fields these enterprises involve are overlapping, but the proportion of resource allocation under the technology innovation layout is different. Samsung Electronics, Co., Ltd. and others avoid International Business Machines Corporation (IBM) 's dominant technology fields, and conduct the technology innovation layout in International Business Machines Corporation's relatively weak fields. These enterprises use this kind of technology innovation layout strategy not only to evade competing with strong enterprises, such as IBM, but also complement each other's technology and create cooperation opportunities with such strong enterprises in the future.

Finally, there are some isolated nodes in the complex network. When the threshold is $|0.4|$, the nodes 4, 33, 51, 57 and 66 in the network are isolated nodes, that the technology innovation layout of the five enterprises has no correlation with other enterprises.

Through the in-depth analysis of their technology innovation fields, it can be found that most patents applied by the five enterprises are the ones that other enterprises seldom involve since 2000-2016. Because of the less overlapping technology innovation field, there is only weak technology proximity with other enterprises. The technology is much different with other enterprises, so they can still be neutral with other enterprises even in the same cloud computing industry. Such enterprises always have specific technologies, they can neither cooperate nor compete with others. Thus maintaining the relationship with such enterprises and never making it become the rival can lay the foundation for future cooperation and innovation.

5. Conclusion

Cloud computing, as another new idea in the computer field, is being paid more and more attention. How to accurately grasp the technology innovation layout of other enterprises, identify potential competitors and find partners have become urgent problems to be solved. Although the research method that analyzed the correlation of technology innovation layout through quantifying the technology proximity had made some achievements, but because the quantitative results were positive, it could not describe the technology complementation and neutrality. For this reason, this paper adopted the weighted bipartite graph projection algorithm and Pearson correlation coefficient algorithm to calculate the correlation coefficient of cloud computing enterprises' technology innovation layout taking full consideration of total enterprise resources and the innovation resource ratio. The research revealed three kinds of correlations of technology innovation layout, and excavated the potential relationship between enterprises comprehensively and intuitively.

References

[1] M. Woerter: *Technology proximity between firms and universities and technology transfer.* The Journal of Technology Transfer *37* (2012), No. 6, 828-866.

[2] K. Hussinger: *On the importance of technological relatedness: SMEs versus large acquisition targets.* Technovation *30* (2010), No. 1, 57-64.

[3] J. Sears, G. Hoetker: *Technological overlap, technological capabilities, and resource recombination in technological acquisitions.* Strategic Management Journal *35* (2014), No. 1, 48-67.

[4] C. S. Curran , J. Leker: *Patent indicators for monitoring convergence–examples from NFF and ICT.* Technological Forecasting and Social Change *78* (2011), No. 2, 256-273.

[5] H. Candelin-Palmqvist, B. Sandberg, U. M. Mylly: *Intellectual property rights in innovation management research: a review.* Technovation *32*(2012), No. 9,502-512.

[6] S. B. Chang: *Using patent analysis to establish technological position: Two different strategic approaches.* Technological Forecasting and Social Change *79* (2012), No. 1,3-15.

[7] J. Yoon, K. Kim: *An analysis of property–function based patent networks for strategic R&D planning in fast-moving industries: the case of silicon-based thin film solar cells.*Expert Systems with Applications *39* (2012), No. 9,7709-7717.

[8] S. C. S. Lo: *A comparative study of linkage indexes: Co-assignee, reciprocal citation, patent coupling and co-patent.*Journal of Library and Information Studies *8* (2010),No. 1, 11-27.

[9] K. Angue, C. Ayerbe, L. Mitkova: *A method using two dimensions of the patent classification for measuring the technological proximity: an application in identifying a potential R&D partner in biotechnology.* The Journal of Technology Transfer *39* (2003), (2014), No. 5,716-747.

[10] Y. Hong, Y. M. Li: *Visualization of technological similarity analysis for enterprises based on patent coupling approach.*Studies in Science of Science *31*, (2013), No. 7,1013-1021.

[11] S. Y. Feng, X. F. Yuan, Y. Xu, J. M. Wang: *Research on patent bibliometrics in cloud-computing field*. Library and Information Servic *57*(2013), No. 9,107-112.

[12] K. Riesen, M. Ferrer: *Predicting the correctness of node assignments in bipartite graph matching*. Pattern Recognition Letters *69*(2016), 8-14.

Received November 16, 2017