Application of Citation Relationship and Co-word Method on Patent Analysis of Cloud PaaS

Jia-Yan Huang [1*] and Pei-Hua, Zhuang

[1]Department of Information Management, National Chin-Yi University of Technology, No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung City 41170, Taiwan (R.O.C.)
[*]Corresponding author

Abstract—Cloud computing is generally considered to be beneficial for the public and private IT organizations because it can substantially reduce the cost of IT systems. Cloud computing technology is not only very important for the development of businesses but also has significant impact on the overall competitiveness of a country. However, there are few patent analyses in this area and many gaps are still open. This research retrieves the mainstream PaaS patents by taking the citing and cited patents into account. Next, we propose co-word analysis, incorporating with the visual methods of social network analysis (SNA), to investigate the R&D themes of PaaS cloud patents. The proposed method for exploring the intellectual structure, as well as the analyzed results for unveiling the development status of cloud computing, can provide valuable references for cloud-related companies to make their R&D management strategy.

Keywords-cloud computing; patent analysis; citation analysis; network analysis; co-word analysis

I. INTRODUCTION

In general, enterprises that own more patents and raise the innovative value of their patents deliver much better returns to their shareholders (Chen and Chang, 2010). In the era of knowledge-based economy, intellectual property has become the tool of competition for enterprises, and there is no exemption for the cloud computing industry. Assessment of the developing trends in the cloud computing industry using patent analysis is essential for enterprises who intend to have prosperous development in this industry.

The National Institute of Standards and Technology (NIST) has defined three service models (known as the SPI model) relating to cloud computing: Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS) (Liu, 2011). SaaS is a software licensing and delivery model in which software is licensed on a subscription basis and typically functions remotely as a Web-based service. It is the application that consumers use, but it does not control the operation system, hardware, or the basic Internet structure of operation. PaaS is a category of cloud computing services that delivers applications over the Internet and host users' hardware and software on the infrastructure of the providers. IaaS is the delivery of computer hardware, such as servers, networking technology, and storage space, as a service. Consumers of IaaS can control the operating system, storage space, deployed applications, and network components (such as firewalls and load balancers), but cannot control the basic structure of cloud.

Due to the space limitations imposed for this conference proceeding, this paper focuses on the study of PaaS.

To avoid duplicating existing research, wasting money, and possibly violating opponents' patents, the strategy of patent application of cloud computing industry should be organized with overall considerations. However, related literature on patent analysis is limited.

Since a specific patent usually has certain connections with other patents, it is more appropriate and easier for analysts to visually investigate the overall relation between patents and the distribution of patent clustering. The emerging methods to discover the patent intelligence, termed patent informatics, use automated tools to fulfill diverse requirements through the techniques of visualization and text mining (Abbas, et al., 2014).

Visualization is an effective way of representing the results of patent analysis compared to other conventional techniques (Shih et al., 2010). Text mining, which has attracted much attention in recent years, includes not only the traditional data mining techniques, but also techniques of other areas such as natural language processing, information extraction, statistics, machine learning, and text analysis. Text mining can help users to locate the necessary features and examine the relationships among a large amount of text data (Tseng et al. 2007).

Considering an increasing number of studies have incorporated text mining and clustering methods with visual methods, this study will utilize text mining-based network analysis and co-word analysis to identify the intellectual structure of cloud patents. Before conducting patent analysis and developing the classification system, the task of patent retrieval must be accomplished in advance. The current study retrieves the patents with citation relationship, including citing patents and cited patents. Based on the retrieved patents, we conduct co-word analysis and social network analysis to reveal the current intellectual structure and competition status of cloud computing industry.

The remainder of this paper is organized as follows. In Section 2, we review the literature related to patent analysis of cloud computing. Section 3 presents the proposed methodology. Section 4 describes the process and results. Finally, we conclude the paper in Section 5.

II. LITERATURE REVIEW

Patent analysis is conducted to transform patent data into usable information through sifting and gathering statistics. It can be applied to different assessment areas, such as competition position, innovation investment, and patent portfolio (Zhang, 2011). Although many big players of the software and communication technology industries are rushing into the market of cloud services, very few analyses have been reported about the patent analsis of cloud patents. Most studies of cloud computing have focused on the technologies development, security, and commercial application (Hudic and Weippl, 2012); few have focused on the issue of patent analysis.

Huang (2016) used text mining as a tool to extract keywords and transfer the original patents data into structured data. Using measures of similarity between patents, the complex relationship among different cloud computing patents was presented as a visual network. With the aid of visual networks and technology centrality indices, he seek to reveal the overall patent structure and identify the key technologies of cloud computing. However, in his study, many unrelated patents were retrieved and most of the patent scattering and disconnected in the visual network diagram, and which causes difficulty in disclosing the status and association relationship between technologies. Moreover, many assignees avoid using specific keywords in the process of patent application, which make it difficult to retrieve the desired patents. As a remedy, the current study focuses on the analysis of patents with citation relationship since related patents can be clustered effectively by using citation analysis.

In bibliometrics, it is common to use citation approach for the assessment of similar patent documents (Weng and Lai, 2009). The citation relationship between patents can be used as the basis for building a technological network. The citation relationship includes backward citation (or citing citation) and forward citation (or cited citation). The information of citing patent and cited patent can be easily obtained from software or search engines, such as Google patent. In general, a patent that cites more prior patents in classes other than those it belongs to is considered to be more radical (Shane, 2001) or higher originality (Kjersten and Laurel, 2005). Contrarily, the more frequent a patent is cited, the higher its influence in a particular technology field can assumed to be. In other words, the more a patent is cited, the patent is considered as having better quality (Lai and Wu, 2005).

Co-citation analysis of bibliometrics has been applied in many specific researches. For example, journal co-citation is of interest to the collection manager concerned with developing core journal lists, and author co-citation analysis has been used in analyzing the intellectual structure of science studies. The study of Stuart and Podoly (1996) is the pioneer of using co-citation approach to patent management. Lai and Wu (2005) used the patent co-citation approach to establish a new patent classification system. The focus of the patent co-citation analysis is on the patents cited, by calculating the frequency of two patents that are co-cited by specific patents. Based respectively on the number of co-citing or co-cited patens, the similarity of two patents can be assessed.

Co-citation is defined as the frequency with which two documents are cited together. In the past, most of literature regarding citation analysis focused on co-citation relationship between patents. Considering that citing patent and cited patent have their respective meaning for patent analysis, this study focuses on the comparison between the intellectual structures of citing patents and cited patents.

III. METHODOLOGY

We conduct co-word and network analysis to explore the intellectual structure of PaaS and the co-competition relations between companies. The framework of this study is shown in Figure 1.

Before conducting patent analysis and developing the classification system, the task of patent retrieval must be accomplished in advance. Considering that cloud computing is inherited from distributed computing, basically, cloud computing patents belong to two IPC classes: G06 (including Electronic Data-Processing, Data Handling System or Data Processing Means, and Information Storage Memory) and H04 (Telecommunication and Communication). However, IPC H04 and/or G06 are vast classes whose meaning is open to interpretation; thus, using the IPC classes as an analytical basis are not encouraged. This study uses an ontology-based compound retrieval policy proposed by Huang (2017) to collect patents in line with the characteristics of the cloud computing industry.

The two steps of traditional co-word algorithm producing the desired networks are: (1) select the link having largest strength as the starting nodes to construct the internal-linked network; (2) Add other related nodes to the network using a breadth-first search on the strength of the links (Delecroix and Epstein, 2004). Considering the problems regarding traditional co-word analysis, which relies on the measurement of association strength, this study employs the co-word analysis method proposed by Huang (2017). The co-word network includes three parts: the internal link network, the external link network, and the extensional link network. One may refer to the work of Huang (2017) for the detailed analysis process.
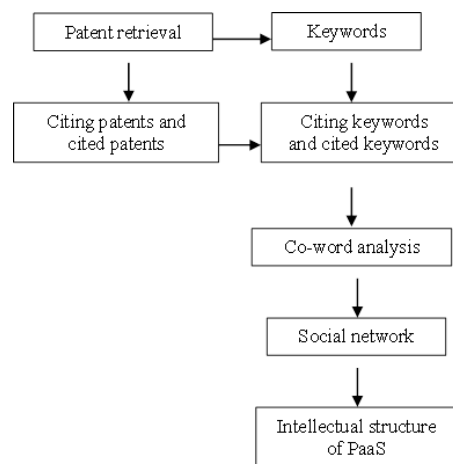

FIGURE I. THE RESEARCH FRAMEWORK OF THIS STUDY

IV. DATA ANALYSIS AND DISCUSSION

A. Retrieval of Patents with Citation Relationship

1)   We Acquired 7413 Patents via the MTrend, an online searching tool developed by Learningtech Corp. Using the procedure proposed by Huang (2017) for patent screening, in the end, we collected 2,020 patents. Among the overall patents, 987 patents were categorized as IaaS, 361 as PaaS, and 672 as SaaS.

2)   By Employing Google Patent, citing patents and cited patents of each retrieved PaaS patent are identified. On the whole, there are 2610 citing patents and 2454 cited patents. Among the 361 PaaS patents, 21 patents are deleted because they have no citing patents and cited patents.

B. Screening of Citing Patents and Cited Patents by Social Network

Based on the results of the retrieved citing patents and cited patents, two citation matrixes are formed. One is the backward citation matrix (or citing citation matrix), and the other is the forward citation matrix (or cited citation matrix). To conduct network analysis, it is necessary to dichotomize the entries of the matrix, which is constructed using the cosine similarity values between nodes. We proceed the network analysis by UCINET6, a social network analysis software, developed by Borgatti, Everett, and Freeman who are from the University of California at Irvine.

The social network of patents with backward citation relationship is shown in Figure 2. There are 183 patents clearly clustered together, meaning that these patents are similar. We therefore focus the investigation of the technologies and companies within this cluster.

Similarly, there are 80 patents clearly clustered together for the patents with forward citation relationship, as shown in Figure 3.

FIGURE II. SOCIAL NETWORK OF PATENTS WITH BACKWARD CITATION RELATIONSHIP
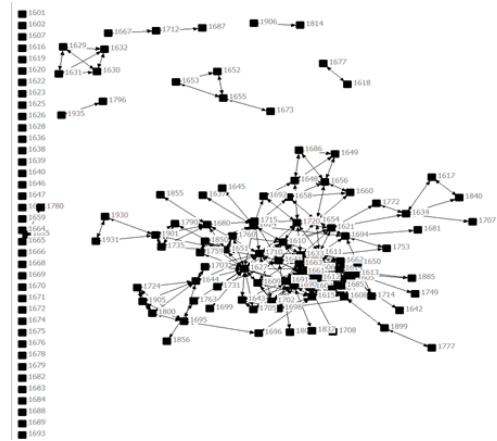

FIGURE III. SOCIAL NETWORK OF PATENTS WITH FORWARD CITATION RELATIONSHIP

C. Co-words Analysis

141 keywords are selected based on the order of TF-IDF of each keyword, which is calculated by the product of term frequency (TF) and inverse document frequency (TF-IDF). One may refer to the study of Trappey et al. (2010) for the formulation of TF-IDF.

We conduct co-word analysis on citing patents and cited patents, respectively. To build the internal link networks, we first construct the patents-terms incidence matrix; next we construct the adjacency matrix by using the cosine similarity between terms as the elements. Then, we compute the degree centrality of terms and rank in descending order.

Only co-words with occurrence frequency higher than 30 are retained. We follow the procedures proposed by Huang (2017) for constructing the external link network and the extensional link network. For citing patents, 41 co-words are found. We use co-word to indicate a technology; hence, five isolated single words are deleted. In the end, there are 36 co-words included in the citing patents. For cited patents, 46 co-words are found. Among these, 9 words that isolated are excluded in the following procedure of building the intellectual structure.

D. The Intellectual Structures

Based on the results of co-words analysis, we can build the adjacency matrix by using the cosine similarity between co-words as the elements; then we construct the intellectual structure by employing the social network analysis on the co-words.

As shown in Figure 4, the intellectual structure of citing patents can be divided into three clusters. Among these, the two clusters on the right contain many technologies related with keyword application, such as: application instance, active application, and application component. While the cluster on the lower right contains many presents many co-words that include the term readable, such as computer readable, readable medium and computer readable medium. The cluster on the left contains many technologies related with virtual, which is a very important technology in cloud computing.

As shown in Figure 5, the intellectual structure of cited patents shows one major cluster on the right. Clearly, the technologies for the cited patents are related with virtual (such as virtual network and virtual image) and storage (such as storage medium, and readable storage). The analysis results are reasonable, since many of the data are stored in the servers of the third party for PaaS service model, the technologies related to storage are therefore of concern to the service providers.
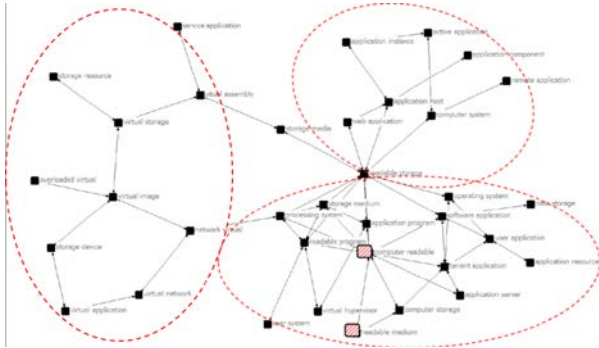

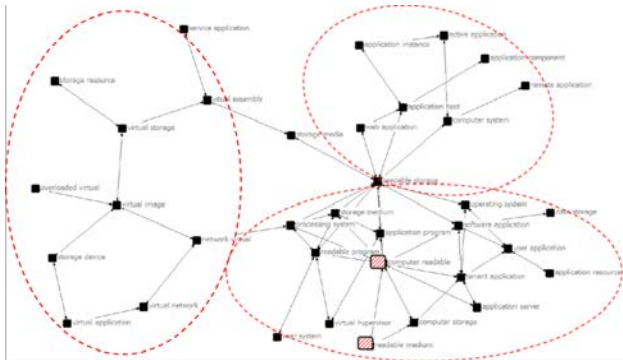FIGURE IV. THE INTELLECTUAL STRUCTURE OF CITING PATENTS

FIGURE V. THE INTELLECTUAL STRUCTURE OF CITED PATENTS

V. CONCLUSION

The major academic contribution of this study is to incorporate co-word algorithm with social network to identify the mainstream technologies and intellectual structure of PaaS patents. In view of the different characteristics between citing patents and cited patents, the current study also attempts to compare the results between these two kinds of patents. In general, a patent that cites more prior patents in classes other than those it belongs to is considered to be more radical or higher originality. Contrarily, for a patent being cited frequently is considered to have better quality or having higher influence in a particular technology.

Based on the analysis results, the intellectual structure of citing patents and cited patents are slightly different. For the citing patents, application is the most important technology, while for cited patents, virtual and storage received more attention. The practical contributions of this study are the disclosure of the competition relationship between PaaS related companies. Overall, both from the analysis results of citing patents and cited patents, Microsoft and IBM are in the lead in PaaS technologies.

REFERENCES

[1] Abbas, A., Zhang, L., Khan, S.U., (2014). A literature review on the state-of-the-art in patent analysis, World Patent Information, 37, 3-13.

[2] Chen, Y.S. and Chang, K.C., (2010). The relationship between a firm's patent quality and its market value - The case of US pharmaceutical industry, Technological Forecasting and Social Change, 77(1), 20-33.

[3] Delecroix, B., & Epstein, R. (2004). Co-word analysis for the non-scientific information example of Reuters Business Briefings. Data Science Journal, 3, 80-87.

[4] Hudic, A., & Weippl, E. (2012). In Private cloud computing: Consolidation, virtualization, and service-oriented infrastructure. 1st Edition by Smoot, S. R., & Tan, N. K. Elsevier.

[5] Huang J.Y., "Patent Network Analysis of Cloud Computing by Text Mining," Journal of Technology, Vol.31, No.2, pp. 127-146, 2016.

[6] Huang J.Y., An analysis of the intellectual structure of the cloud patents of SaaS, Technology Analysis & Strategic Management, Vol.29, No.8, pp.917–931, 2017.

[7] Kjersten, B. W. and Laurel. S. D., Quality versus Quantity: Women's Patenting in the Life Sciences, the Society for the Social Study of Science, 2005.

[8] Lai, K. K., & Wu, S. J. (2005). Using the patent co-citation approach to establish a new patent classification system. Information processing & management, 41(2), 313-330.

[9] Liu, F., Tong, J., Mao, J., Bohn, R., Messina, J., Badger, L., and Leaf, D. (2011). NIST cloud computing reference architecture. NIST special publication 500-293, pp1-85.

[10] Shane, S. (2001). Technological opportunities and new firm creation. Management science, 47(2), 205-220.

[11] Shih, M.J., Liu, D.R., & Hsu, M.L. (2010). Discovering competitive intelligence by mining changes in patent trends. Expert Systems with Applications, 37(4), 2882-2890.

[12] Stuart, T. B., & Podoly, J. M. (1996). Local search and the evolution of technological capabilities. Strategic Management Journal, 17, 21–28.

[13] Trappey, C.V., Trappey, A.J.C., & Wu, C.Y., (2010). Clustering patents using non-exhaustive overlaps. Journal of Systems Science and Systems Engineering, 19(2), 162-181.

[14] Tseng, Y. H., Lin, C. J. and Lin, Y. I., (2007). Text Mining Techniques for Patent Analysis, Information Processing and Management, 43(5), 1216-1247.

[15] Weng, C. S. and Lai K. K., (2009), On the Technological Isomorphism of Insurance Business Method Patents-The Perspective of Social Network Analysis, Journal of Management, 26(5), 485 – 506.

[16] Zhang, Y., (2011). Analysis and evaluation of enterprise innovation ability conversions, International Journal of Innovative Management, Information & Production, 2(2), 39-46.