International Journal of Business and Economic Affairs (IJBEA)
5(1), 1-8 (2020)
DOI: 10.24088/IJBEA-2020-51001
ISSN: 2519-9986

Discovering Emerging Financial Technological Chances of Investment Management in China via Patent Data

KUNG WANCHIA [1*], JIA YUFEI [2], YU HSINCHUN [3]

[1,2,3] Tunghai University, Taichung City, Taiwan

Abstract: This paper aims to explore and identify commercial opportunities about the financial technologies used in investment management in China. We focus on a method using KeyGraph, a text data mining tool for extraction and visualization of preference from information to assist the process of chance discovery, to reveal the structure of the patents which have CPC G06Q40/06, the classification label that covers technologies related to financial investment. Consequently, through conducting both cluster and association analysis on the patent data, Y10S707/99935(Query augmenting and refining) has been found as a chance that connects advanced investment analysis system and the basic information management system. In addition, the result shows that the presented method is promising to generate new opportunities from the chance nodes obtained from the KeyGraph. Patents are a valuable source of information in technical innovation, market direction, and competitive intelligence. Discovering emerging chances from the patent data helps to improve planning for business decisions in further investments accordingly.

Keywords: Chance discovery, KeyGraph, Financial technology, FinTech patent, Text data mining

Received: 04 October 2019 / Accepted: 3 January 2020 / Published: 27 February 2020

INTRODUCTION

In business competition, almost all important inventions or technologies apply for patents. Besides, patents tend to emerge before technical products or services appear on the market. Hence, if companies make full use of patents ahead of the market and see the future trend of competition, they will be able to make early plans against market changes, avoid potential disadvantages and even gain advantages. In recent years, with the development of various technologies, Finance Technology (Fintech) has achieved big leaps, induced great changes in the financial market, institutions and service provision. In the existing market of finance, excellent finance technology becomes the source of profit basis and competitiveness for companies.

Within the framework of Cooperative Patent Classification (CPC), there are six major patent classifications concerning financial technology: Payment (CPC: G06Q/20), Investment (CPC: G06Q40/06), Banking (CPC: G06Q40/02), Exchange (CPC: G06Q40/04), Insurance (CPC: G06Q40/08) and Tax strategies (CPC: G06Q40/10). Among them, Investment (CPC: G06Q40/06) achieved remarkable progress owing to technical improvements like cloud computing, big data and artificial intelligence. Regardless of proposing auxiliary suggestions or seeking investment strategies, the concept of robot financing has given people infinite chances of imagination (Layyinaturrobaniyah, Masyita, & Sekartadjie, 2016).

Therefore, the purpose of this study is to seek potential business opportunities relating to finance technology and Investment (CPC: G06Q40/06). Chance Discovery theory (Ohsawa, 2003, 2005, 2006) and KeyGraph technology (Kushima et al., 2017; Matsumura et al., 2002; Ohsawa, Benson, & Yachida, 1998; Sayyadi & Raschid, 2013), proposed by Ohsawa in the field of Text data mining, are used to analyze relevant patents released in Investment (CPC: G06Q40/06) by the end of 2019, and demonstrate the analysis results in visual graphs that are easy for comprehension. Chance Discovery theory and KeyGraph

*Corresponding author: Kung Wanchia
†Email: wanchiakung@gmail.com

are different from average Text data mining technologies, which often aim to seek high frequency targets. The target chance of chance seeking is the project with low frequency. This project gives rise to the connection between the two content entities. That is to say, this chance serves as the bridge between the two content entities and creates decisive impacts as well (S. Lee, soo Kim, Park, & Kim, 2016). As far as this study is concerned, this chance may become the focus of patents in finance technology. For this reason, early identification of this chance is likely to help companies achieve new development or avoid risks.

LITERATURE REVIEW

Chance discovery

Chance discovery is a text data mining method put forward by Ohsawa and Fukuda (2002). Data mining is the procedure that aims to find valuable information from abundant disorderly data. Most data mining methods strive to obtain meaningful patterns or high frequency items, which are extremely important and valuable. However, this type of information is relatively apparent, as it is essential to every competitor in business competition. That is to say, such information can guarantee that companies will not be eliminated in business competition. But it can not enable them to outperform their competitors. All it is able to guarantee is equal chances. Chance discovery aims at obtaining the cross point of meaningful scenarios. The cross point is not a high-frequency item in the whole data, but it may involve important information value. The cross point is an inconspicuous detail compared with apparent information. It is because of this unique trait that cross point is important and yet easy to be overlooked, which is more similar to chances. To put it in another way, although chances do not take place frequently, they are the critical events with a significant influence on decision making (Ohsawa, 2003).

When it comes to the process of chance mining and result application, human plays quite an important role. As is shown in the chance discoverys procedure in Figure 1, human and data mining system shows mutual improvement. This procedure starts when people focus on a certain even or the development of a certain field. They collect data from related fields of the new focus, and conduct literal data mining in the dataset input information system. Then, the results are demonstrated in clear and understandable ways to those involved in research and discussion, such as KeyGraph used in this paper. Through observing and understanding the dynamic changes of the external world, people discuss the rare events or status concerning the new impacts of the results in common fields. Next, after explanation or reassessment, the process of data gathering or preparation, thereby generating new data. Lastly, literal data mining is carried out to gain new results, which are further discussed, assessed and understood by people. The above cycle is repeatedly conducted to combine human intelligence with data analysis. Chances are discovered in the process, so as to make decisions and generate the most valuable result.

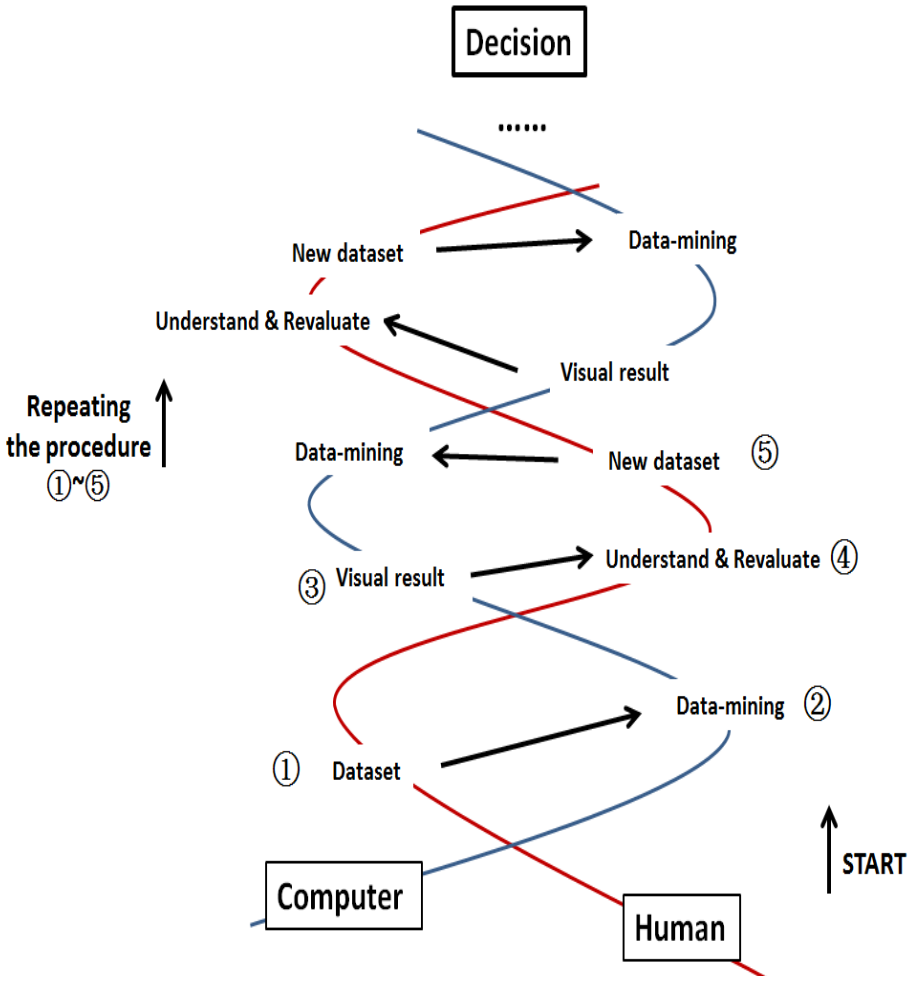

Figure 1. The chance discoverys procedure

Chances have always existed, but they are only meaningful when they are discovered. The purpose of chance discovery is to perceive and capture events or situations that are often ignored as noise, so as to interpret and utilize the scare events and results. For this reason, chance discovery technology is not only used by computer data scientists and engineers, as well as researchers in professional fields like psychologists, philosophers, economists and sociologists, etc. This technology is also adopted in relevant research and studies, including identifying the potential domestic consumption needs to offer suggestions on improving the quality of meals (Ohsawa & Fukuda, 2002); analyzing seismic records to predict risks of earthquakes (Ohsawa, 2002); getting information from recipe-sharing websites, interpreting user preferences and seeking new business opportunities (Emoto, 2015; Shin & Seo, 2017); formulating more accurate and personalized recommendation system based on chance discovery theory (Kong et al., 2018); conducting chance discovery process according to the content of online discussion etc.(Seo, Iwase, & Takama, 2006).

KeyGraph

KeyGraph is one of the most important analytical tools for chance discovery. Initially, it was a visualization tool for the generalization index of text files, which was used to gather the implicitly critical information and seek the causal relationship from data. How to clearly convey the important events to decision makers and how decision makers should formulate policies based on such information are also critical issues. According to some scholars, instead of texts, graphs are more able to depict and convey information, thereby assisting the process of decision making. Hwang found that in comparison with forms, visualization graphs can better convey in spite of time pressure (Hwang, 1995). KeyGraph was proposed by Ohsawa for the first time in 1998, which is a tool using visualization graphs to convey information to decision makers. KeyGraph can demonstrate images for the procedure of chance discovery.

In the past, KeyGraph was used to visualize the comments of blogs, thereby identifying the core contents (Tsuda & Thawonmas, 2005); as for the care life log data generated by the long-term health care and nursing services, patients degrees of health and caring are analyzed (Kushima et al., 2017); the data of mobile context sharing system are analyzed to screen those who focus on internal communication and others that emphasize socializing (Khan, Hussain, Shahbaz, Yang, & Jiao, 2020; M.-C. Lee, Lee, & Cho, 2013).

Besides, KeyGraph offers the interpretation based on graphs. Figure 2 shows the icon examples analyzed by KeyGraph, including the following basic elements:

- Black nodes indicate items that frequently occur in a data set.
- Red nodes indicate the items that occur less frequently overall but frequently occur with black nodes in a data set.
- Green circles indicate the items that have the strongest effect to connect multiple clusters, which can be considered as keywords.
- Links indicate that the connected item pair cooccurs frequently in a data set. Solid lines form a foundation, which dotted lines connect foundations.

There are three types of structural relationships between basic elements, which are usually named with construction structures for better understanding. The descriptions are shown in the following:

1. Foundations (Clusters) - Subgraphs of highly associated and frequent terms that represent basic concepts in the data. Foundations consist of black nodes and are connected with solid lines. Each of the foundations are constructed by the co-occurrence among the terms in the document.
2. Roofs (Chances) - Data with nonfrequent appearance but high relevancy with foundations. They are red dots with green circles.
3. Columns - Connect roofs and foundations to seek key words. Columns are shown with dotted lines. Although foundation demonstrates contents known to the public, the contents of connection between columns are not known.

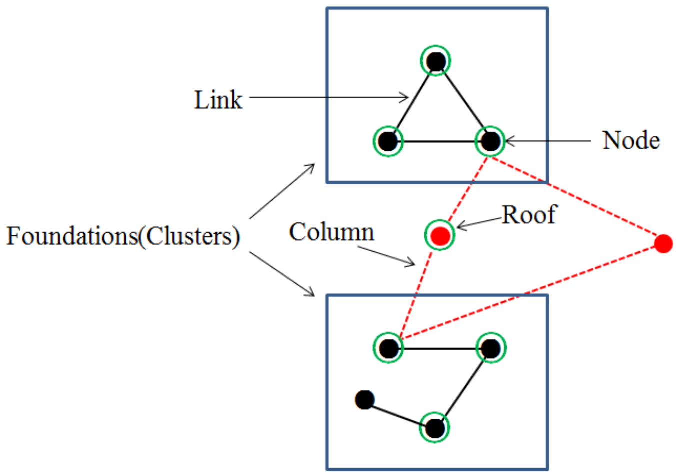

Figure 2. An example diagram of KeyGraphs composition

RESEARCH METHOD

Through the patent retrieval entry of National Intellectual Property Administration, PRC (http://pss-system.cnipa.gov.cn/sipopublicsearch/portal/uiIndex.shtml), the study gathered and arranged all the patents in the field of finance technology investment (CPC: G06Q40/06) within CPC classification system since the establishment of patent system to Dec.31, 2019. A total number of 1892 patents were gathered, which are concerned with 736 categories of patented technologies.

In this study, Polaris (http://www.panda.sys.t.u-tokyo.ac.jp/KeyGraph/), a KeyGraph tool provided by Ohsawa, is used for chance discovery in patents relating to finance technology investment. Parameter settings of this experiment include 30 black nodes, 32 black links, 25 red nodes and 10 green nodes. Jaccard coefficient is used to measure data items. As for the application of data mining, Jaccard coefficient is used to compare the similarities and differences between two limited samples. They are defined as the frequency of the intersection sets of two data items X and Y divided by the frequency of connex sets, which is expressed as:

$$Jaccard\,(X,Y) = \frac{P(XandY)}{P(XorY)} \tag{1}$$

Greater values of Jaccard demonstrate higher levels of similarity. When X and Y are void, Jaccard (X, Y) = 1. KeyGraph parameter setting does not have regular rules. Excessive black nodes may lead to multifarious and disorderly information. The number of black links decides the level of strong correlation and clusters. When it comes to Jaccard coefficient calculating the data items, if nodes frequency does not show any obvious difference in data items, Jaccard coefficient produces the optimum effects. In this study, many items have approximate frequencies. As many as 75 items have frequency levels ranging from 13 to 20. For this reason, Jaccard coefficient is chosen as the parameter for calculating the strength of the connection.

RESEARCH RESULTS AND DISCUSSIONS

Arranged patent classification data are input to output KeyGraph, such as Figure 3. What should be noted is that the preset parameter quantity is the target value. In fact, as multiple items share the same calculated values, the final nodes and the number of tie links may be different from set values. The results of this analysis include 29 black nodes, 32 black links, 1 red node and 10 green nodes.

According to the analysis results, G06Q40/04 and G06Q40/06 are regarded as cluster 1, whose frequencies of occurrence are 433 and 1892. G06Q40/00, G06Q40/02, G06Q30/0283, G06Q30/0601, G06F16/9535, G06F3/04883 and Y10S707/99933 are regarded as cluster 2, whose frequencies of occurrence are 91, 91, 26, 32, 28, 19 and 18. Y10S707/99935 is roof 1 that combines two clusters, whose frequency of occurrence is 16. Frequencies of occurrence in key nodes (with green circled) are shown in Table 1.

Definitions of patented technologies of cluster 1, cluster 2 and roof 1 are shown in Table 2 and Table 3. Based on more detailed and further analysis, cluster 1 is the analytic strategy system of trade and investment. Cluster 2 is the management system widely used in finance. Roof 1 is the query augmenting and refining technology.

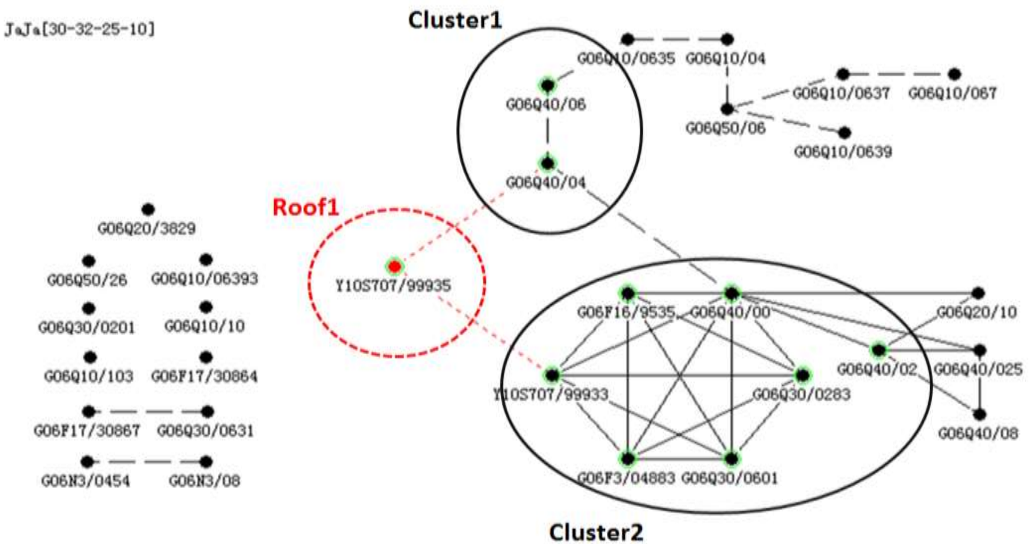

Figure 3. The KeyGraph for the patents involve G06Q40/06

Table 1: The frequencies of the Keywords

Node	CPC	Frequency of occurrence
Black-green	G06Q40/06	1892
	G06Q40/04	433
	G06Q40/00	91
	G06Q40/02	91
	G06Q30/0601	32
	G06F16/9535	28
	G06Q30/0283	26
	G06F3/04883	19
	Y10S707/99933	18
Red-green	Y10S707/99935	16

Table 2: The patents scheme within clusters

Cluster	CPC	Content of patent technology
Cluster 1	G06Q40/04	Exchange, e.g., stocks, commodities, derivatives or currency exchange
	G06Q40/06	Investment, e.g., financial instruments, portfolio management or fund management
Cluster 2	G06Q40/00	Finance; Insurance; Tax strategies; Processing of corporate or income taxes
	G06Q40/02	Banking, e.g. interest calculation, credit approval, mortgages, home banking or online banking
	G06Q30/0283	Price estimation or determination
	G06Q30/0601	Electronic shopping
	G06F16/9535	Search customisation based on user profiles and personalisation
	G06F3/04883	for entering handwritten data, e.g. gestures, text
	Y10S707/99933	Query processing, i.e., searching

Table 3: The patent scheme of the roof node

Roof	CPC	Content of Patent Technology
Roof 1	Y10S707/99935	Query augmenting and refining, e.g., inexact access

It means the analytic strategy system transaction and investment and management system used in finance are mutually independent. In order to integrate them into a more comprehensive and extensive system, the optimized search system is an essential and critical factor, i.e., chance. Practical observations also support this discovery. Financial investment should be promptly adjusted according to social affairs. The most critical way to learn about social changes is the big data analysis. However, people from different cultural groups may use different texts to describe the same situation, making it more difficult to conduct data analysis. Augmenting query technology is one of the possible solutions, which can identify synonyms or attitudes of positive and negative words. With the help of augmenting query technology, transaction and investment analysis system can be widely utilized in finance management, so as to input various big data to generate realtime strategies and evolve into more advanced systems.

CONCLUSION AND RECOMMENDATIONS

Chance discovery aims to help people focus on rare but important information, so as to seek the chances and risks that may be ignored. To seek business opportunities in finance technology of investment strategies, on the basis of KeyGraph, this study carries out chance discovery in patents relating to finance technology of investment by the end of 2019. It finds out Y10S707/99935(Query augmenting and refining) is the chance that connects the analytical strategy system of transaction and investment with the management system of finance. Through increasing the input in research and development of query augmenting and refining technology, it is feasible to enhance the efficiency and practicability of the analytical strategy system of transaction and investment and the management system widely used in finance, thereby enhancing the corporate competitiveness and differentiation.

This study is the first to use KeyGraph to analyze the portfolio of finance technology patents. The methods and results may serve as references for subsequent discussion and research on chances and tendencies of finance technology development. It is suggested conducting research on patent materials in other countries besides China or fields of high economic value and intense competition.

REFERENCES

Emoto, M. (2015). Extraction of preference of recipe providers and users on recipe-sharing websites. In *International Conference on Data Mining Workshop (ICDMW),* Atlantic City, NJ.

Hwang, M. I. (1995). The effectiveness of graphic and tabular presentation under time pressure and task complexity. *Information Resources Management Journal*, *8*(3), 25–31. doi:https://doi.org/10.4018/irmj.1995070103

Khan, Z., Hussain, M., Shahbaz, M., Yang, S., & Jiao, Z. (2020). Natural resource abundance, technological innovation, and human capital nexus with financial development: a case study of china. *Resources Policy*, *65*, 101585. doi:https://doi.org/10.1016/j.resourpol.2020.101585

Kong, D., Fan, Y., Du, Y., Hu, S., Liu, Y., & Li, Q. (2018). Personalized recommendation algorithm based on the chance discovery in social network services. In *International Conference on Cloud Computing and Intelligence Systems (CCIS),* Nanjing, China.

Kushima, M., Araki, K., Yamazaki, T., Araki, S., Ogawa, T., & Sonehara, N. (2017). Text data mining of care life log by the level of care required using keygraph. In *Proceedings of the International MultiConference of Engineers and Computer Scientists,* Kowloon, Hong Kong.

Layyinaturrobaniyah, Masyita, D., & Sekartadjie, G. (2016). Fundamental and technical analyses for stock investment decision making. *Journal of Administrative and Business Studies*, *2*(1), 1-17. doi:https://doi.org/10.20474/jabs-2.1.1

Lee, M.-C., Lee, Y.-S., & Cho, S.-B. (2013). KeyGraph-based social network generation for mobile context sharing. In *International Conference on Green Computing and Communications,Internet of Things, Cyber, Physical and Social Computing,* Zhejiang, China.

Lee, S., soo Kim, M., Park, Y., & Kim, C. (2016). Identification of a technological chance in product-service system using KeyGraph and text mining on business method patents. *International Journal of Technology Management*, *70*(4), 239. doi:https://doi.org/10.1504/ijtm.2016.075884

Matsumura, Y., Nakano, H., Kusuoka, H., Park, K., Matsuoka, M., Oshima, H., ... Takeda, H. (2002). Clinic hospital cooperation system based on the network type electronic patient record. *Japan Association for Medical Informatics*, *22*(1), 19–26.

Ohsawa, Y. (2002). KeyGraph as risk explorer in earthquake - sequence. *Journal of Contingencies and Crisis Management*, *10*(3), 119–128. doi:https://doi.org/10.1111/1468-5973.00188

Ohsawa, Y. (2003). KeyGraph: Visualized structure among event clusters. In *Chance discovery* (pp. 262–275). Heidelberg, Germany: Springer.

Ohsawa, Y. (2005, nov). Data crystallization: Chance discovery extended for dealing with unobservable events. *New Mathematics and Natural Computation*, *01*(03), 373–392. doi:https://doi.org/10.1142/s1793005705000226

Ohsawa, Y. (2006). Chance discovery: The current states of art. In *Chance discoveries in real world decision making* (pp. 3–20). Berlin, Heidelberg: Springer.

Ohsawa, Y., Benson, N., & Yachida, M. (1998). KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor. In *Proceedings of International Forum on Research and Technology Advances in Digital Libraries - ADL98,* Maui, HI. Retrieved from `https://doi.org/10.1109%2Fadl.1998.670375` doi:10.1109/adl.1998.670375

Ohsawa, Y., & Fukuda, H. (2002). Chance discovery by stimulated groups of people. Application to understanding consumption of rare food. *Journal of Contingencies and Crisis Management*, *10*(3), 129–138. doi:https://doi.org/10.1111/1468-5973.00189

Sayyadi, H., & Raschid, L. (2013). A graph analytical approach for topic detection. *ACM Transactions on Internet Technology (TOIT)*, *13*(2), 1–23.

Seo, Y., Iwase, Y., & Takama, Y. (2006). KeyGraph-based BBS for online chance discovery. In *International Conference on Systems, Man and Cybernetics,* Taipei, Taiwan.

Shin, S. J., & Seo, W. (2017). Identifying new technology areas based on firms internal capabilities. *Journal of Administrative and Business Studies*, *3*(3), 114-121. doi:https://doi.org/10.20474/jabs-3.3.1

Tsuda, K., & Thawonmas, R. (2005). Keygraph for visualization of discussions in comments of a blog entry with comment scores. *WESE Trans. Computers*, *12*(5), 794-801.