Extracting Key Technology Using Advanced Fuzzy Clustering

Sunghae Jun* and Seung-Joo Lee
Department of Statistics, Cheongju University, Chungbuk, Korea
shjun@cju.ac.kr, access@cju.ac.kr
*Corresponding Author: Sunghae Jun (shjun@cju.ac.kr)

Abstract
Most companies that have developed their own technology have registered and patented the results. As a result, patent and technology management (PTM) has become important to companies needing to improve their market competitiveness. Using the PTM process, a company can develop a new product that is competitive in their market. First, we need to know the key technology (KT) a company uses to develop new technology. In this paper, we propose a method of extracting the KT for new product development and effective PTM. The proposed method uses advanced fuzzy clustering, which groups patents into clusters according to their technological difference. To verify the performance of our method, we perform a case study using all patents applied for by the company Adobe Systems Incorporated.

Keywords: Key technology, advanced fuzzy clustering, technology management, patent analysis

1. Introduction
Technology has become increasingly important to both companies and countries [1, 2]. In the information and communications technology (ICT) sector, the significance of technology is even greater than others. Most ICT companies that have developed their own technology have registered the result as a patent [3]. A patent is a type of intellectual property (IP) that includes exclusive rights to a technological invention [4]. Most companies apply for patents because a patent is a measure of technological competition. As a result, patent and technology management (PTM) is a key factor in increasing the market competitiveness of a company [1]. Using the PTM process, a company can develop a new product that is competitive in their market. First, we need to know the key technology (KT) a company uses to develop new technology. In this paper, KT is defined as fundamental technology contributing to PTM within a company, including new product development and R&D planning. This research proposes a method of extracting the KT within a company for effective PTM. The proposed method uses advanced fuzzy clustering, which groups patents into clusters according to their technological difference. To verify the performance of our method, we perform a case study using all patents applied for by the company Adobe Systems Incorporated (henceforth, Adobe).

2. Patent and Technology Management
PTM is the management of patent and technology for R&D policy, new product design, new market development, identifying breakthrough technologies, and
technological innovation [1]. The interest in PTM is greater now than ever before. PTM consists of planning, designing, operating, and forecasting of technology based on patents. In the PTM process, finding KT is important to a company’s competitiveness, because the KT performs basic functions in new product development. There are two approaches to performing PTM, namely qualitative and quantitative methods. The qualitative approach is based on the prior knowledge of domain experts. The Delphi method is an example of a qualitative approach [5, 6]. The results of using a qualitative approach are subjective, as they depend on experts’ experience. Depending on the circumstances, these results may fluctuate widely. The quantitative approach is based on the analytical methods of statistics and machine learning and uses objective data from patent documents [7, 8]. This approach tried to overcome the subjectivity problem of the qualitative approach, but also had limitations. For example, an expert was still required to explain the results of a quantitative method. In this paper, we propose a more advanced quantitative method for PTM.

3. Advanced Fuzzy Clustering to Find Key Technology

In this paper, we propose advanced fuzzy clustering as a patent clustering technique. This clustering model can establish the KT used by a company. In crisp clustering, each data point is assigned to only one cluster [9]. However, using fuzzy clustering, a data point may belong to several clusters simultaneously [9]. In this paper, the patent data set $P$ is represented as follows:

$$P = \{p_1, p_2, \ldots, p_n\}$$

Where, $p_i$ is the $i$th retrieved patent. If $k$ shows the number of clusters, we can define a $k \times n$ partition matrix, $F=(f_{ij})$, as follows:

$$0 \leq f_{ij} \leq 1, \quad i=1,2,\ldots,k, \quad j=1,2,\ldots,n$$

In addition, the fuzzy partition matrix satisfies the following conditions:

$$\sum_{i=1}^{k} f_{ij} = 1, \quad \text{and} \quad \sum_{j=1}^{n} f_{ij} > 0$$

We use $f_{ij}$ as a membership value of patent $j$ in the $i$th cluster. In this paper, we use the International Patent Classification (IPC) code as input data for fuzzy clustering. The IPC code is a hierarchical structure of patent technology [4, 10]. In addition, the IPC codes are used as efficient data for patent analysis [11, 12]. Therefore, our structured data uses IPC codes. Figure 1 shows the structured data for patent clustering.

![Figure 1. Structured Data Set](image)

In Figure 1, the rows and columns represent patent and IPC codes, respectively. The element $freq_{ij}$ shows the frequency of IPC code $j$ in the $i$th patent. The IPC codes are
used to cluster variables in our fuzzy approach. In this paper, we obtain the clustering result shown in Figure 2.

![Figure 2. Fuzzy Clustering Result](image)

This figure shows three overlapping clusters. Here, the KT is defined as the patents in the area where the three clusters overlap (diagonal line shading):

$$(\text{Cluster A}) \cap (\text{Cluster B}) \cap (\text{Cluster C})$$

Using this characteristic of fuzzy clustering, we are able to find the patent area with which to define the KT of the target technology. The next figure shows the proposed patent clustering process for extracting the KT.

![Figure 3. Process for Extracting Key Technology](image)

After deciding the target technology, we retrieve the patent documents related to the target technology using a keywords equation. These documents are not suitable for fuzzy clustering because they are not structured. Therefore, we use IPC codes to obtain structured data for our clustering. Using the structured IPC code data, we perform the fuzzy clustering and select the overlapped patents in the clustering result. Finally, we define the KT using the titles and abstracts of the overlapped patent documents. The KT information can be used for new product development and R&D planning. In addition, our results can contribute to diverse fields of technology management.
4. Experiment and Results

We performed a case study to verify the performance of our proposed approach. We used the patent documents for the company Adobe from the Korea Intellectual Property Rights Information Service (KIPRIS) [13]. We searched for the patents assigned by Adobe, obtaining 638 patent documents. Of these, 600 were valid patents that included IPC codes, title, and abstract, which we used as input data. Figure 4 shows the trends of the IPC codes and assigned patents by year.

![Figure 4. Trends of the Numbers of IPC Codes and Patents](image)

Adobe applied for its first patent in 1989. The highest number of patents occurred in 2007. However, the number of patents has decreased in recent years. We knew that the trend of IPC codes was similar to the number of applied patents. Figure 5 shows the number of IPC codes and patents by nation.

![Figure 5. Numbers of IPC Code and Patents by Nation](image)

Here, CA, CN, DE, EP, GB, and WO represent the nations of Canada, China, Germany, Europe, the United Kingdom, and the Patent Cooperation Treaty (PCT). The complete list of IPC codes is as follows: A63F, B41B, B41F, B41J, E04C, E04G, G01D, G02B, G03B, G03F, G03G, G05B, G06F, G06K, G06Q, G06T, G09G, G10L, G11B, H03M, H04B, H04L, H04N, and H04W. From these results, we knew that most patents of Adobe applied to Europe and International areas. To obtain more detailed information, we found the top-ranked IPC codes as follows:
Based on these results, we knew most patents contained the IPC codes G06F and G06T. Therefore, we concluded that the technologies based on these codes were important to Adobe. We used the top five IPC codes for our fuzzy clustering, as shown in the following graph.

We know that the annual trend of G06T is different to other IPC codes. Most development for the technology of G06T occurred between the late 1990s and early 2000s, and development in this area has since decreased. In contrast, the technologies of the other four IPC codes have seen substantial development in recent years. Table 1 shows the representative technologies of the top five IPC codes [14].

<table>
<thead>
<tr>
<th>IPC code</th>
<th>Defined technology</th>
</tr>
</thead>
<tbody>
<tr>
<td>G06F</td>
<td>Electric digital data processing</td>
</tr>
<tr>
<td>G06T</td>
<td>General image data processing or generation</td>
</tr>
<tr>
<td>H04N</td>
<td>Pictorial communication, e.g., television</td>
</tr>
<tr>
<td>G06K</td>
<td>Recognition of data; presentation of data; record carriers; handling record carriers</td>
</tr>
<tr>
<td>G09G</td>
<td>Arrangements or circuits for control of indicating devices using static means to present variable information</td>
</tr>
</tbody>
</table>
Most of Adobe’s patents were based on the technologies of \textit{G06F} and \textit{G06T}, in other words, “electric digital data processing” and “general image data processing or generation.” These then are the representative technologies of Adobe. In addition, the technology of \textit{H04L} was ranked sixth in the IPC codes. This technology is “transmission of digital information and telegraphic communication.” Therefore, we concluded that the areas of processing and generating technologies of electric digital and image data are important to Adobe. Figure 8 shows a box-plot of the number of IPC codes by nation.

![Figure 8. Distributions of IPC Codes by Nation](image)

The dispersion of \textit{DE} IPC codes is the largest, with the largest distribution occurring in Germany. Next, we performed patent clustering using fuzzy clustering. First, we determined that there were two clusters using the criterion from the hierarchical approach. Figure 9 shows our clustering result.

![Figure 9. Patent Clustering Result](image)

In this paper, we selected the patents in the overlapping area of the two clusters. These patents were used to define the KT for Adobe. Next, we extracted keywords from the titles and abstracts of these patents. Table 2 shows the patent numbers and keywords extracted from the titles and abstracts of the patents.
Table 2. Patent Numbers and their Extracted Keywords

<table>
<thead>
<tr>
<th>Patent numbers</th>
<th>Extracted keywords</th>
</tr>
</thead>
</table>

From the text database of titles and abstracts, we removed the common terms, such as “is” and “that.” We then selected the nine keywords with the highest frequencies. Using these keywords, we defined the KT for Adobe, namely the technology of displaying, identifying, and rendering digital color text and image data. Therefore, we can suggest that Adobe develops new products using this KT.

5. Conclusion

For PTM, we proposed an advanced fuzzy clustering method for finding the KT for a company using the patent documents filed by that company. The KT is defined based on the keywords extracted from the patents and that lie within the overlapping area of the clusters, because these patents could represent the technology common to all development within the company. To verify the performance of our work, we carried out a case study using the patent data of Adobe Systems Incorporated, an ICT company. Our analysis suggested two technological clusters. We performed fuzzy clustering for all patent documents filed by Adobe. We then found the patents included in the two clusters and extracted the keywords from the titles and abstracts of these patents. Using these keywords, we defined the KT for Adobe. The KT information of a company offers diverse benefits to that company, including new product planning and R&D policy. In future work, we will research more elaborate approaches to extracting KT information from patent data.

References


Authors

**Sunghae Jun** is an Associate Professor in the Department of Statistics, Cheongju University, Chungbuk, Korea. He received B.S., M.S., and PhD degrees from the Department of Statistics, Inha University, Incheon, Korea in 1993, 1996, and 2001, respectively. He also received a PhD degree from the Department of Computer Science, Sogang University, Seoul, Korea in 2007, and a Doctorate in Information Management Engineering from Korea University, Seoul, Korea in 2013. He was a visiting scholar at the Department of Statistics, Oklahoma State University, Oklahoma, USA during the period 2009–2010. His current research interests include data mining and technology forecasting.

**Seung-Joo Lee** is a Professor in the Department of Statistics, Cheongju University, Chungbuk, Korea. He received a B.S. degree from the Department of Applied Statistics, Cheongju University, Chungbuk, Korea in 1985, and M.S. and PhD degrees from the Department of Statistics, Dongguk University, Seoul, Korea in 1987 and 1994, respectively. His current research interests include Bayesian statistics and technology forecasting.