

An Analysis of Soccer-Related Patents

Kyung-Hoon Park*

Department of Sports Coaching, Jeonju University, Korea;
pkh0033@hanmail.net

Abstract

Background/Objectives: This research was conducted to analyze domestic soccer-related patents and to identify trends in patents for domestic soccer and make predictions about domestic soccer technology using statistical methods and data mining methods. **Methods/Statistical Analysis:** The analysis object data used by this study targeted domestic patents and utilities as 'soccer'-related patents. The period for the search was not subjected to any specific restriction. All patents registered, disclosed and expired were targeted in the process of collection and 185 patents ranging from year 1973 to 2015 based on the year of patent application were selected as analysis objects. Cluster analysis, association technology extraction and technology network analysis were used for statistical analysis. **Findings:** As a result, in soccer-related technologies, the technologies related to soccer footwear or robots are predicted as becoming significant technologies in the realm of off-line services and soccer game-related technologies are predicted as becoming significant technologies in the realm of on-line services. **Application/Improvements:** This study identified the relationships among the technologies of soccer-related industries and central technology clusters and patents. It is thought that the findings of this study can be used as basic data to make predictions about the technologies of soccer-related industries.

Keywords: Association Technology, Data Mining, Soccer, Patent, Social Network Analysis

1. Introduction

In modern society, sports are becoming more scientific and the importance of sports equipment is emerging with the importance of physical, psychological, strategic and technological factors¹. According to reports, in the case of skating, the athletic performance of ice sports actually increased by about 5% because of the appearance of klapskate, compared to previous years² and also the athletic performance of splints increased because of the use of elastic shoe soles. Sports equipment to improve athletic performance has been developed in almost all types of sports. Naturally, in the case of soccer, soccer equipment such as balls, shoes, goal posts, kits and gloves are being newly developed. The phenomenon of sports becoming scientific greatly affects the growth of the related industries. Growth is occurring in the industries of the broadcast relay equipment for live broadcasting and the wearable sports devices to check the conditions of athletes and to improve athletic performance³. Table soccer game, which is among one of the most popular games, is a good example of a typical new technology industry

related to soccer that has appeared with the advent of the Information Age⁴. As sports industries develop, the related promising technologies are being developed. Promising technologies mean those technologies that represent industrial competitiveness and which have advantages in terms of performance, quality, efficiency, etc. Promising technologies in the sports industries mean new technologies applied to sports equipment, service facilities, etc. As the industries related to soccer are continuously growing, it is very meaningful to predict new promising technologies by systematically analyzing an ever-increasing flow of relevant technological information.

The existing methods of technology prediction are classified into two main categories⁵. First is the quantitative analysis method such as the Delphi method⁶ and Tree method⁷ to observe technology trends through the meeting and coordination of views among experts. Second is the qualitative method such as trend analysis⁸ and patent analysis⁹⁻¹¹. Among these, patent analysis is an approach to derive information useful for a specific technology by collecting and analyzing patent

* Author for correspondence

documents including the results of the development of the technology and to effectively carry out technology predictions using the results of the patent analysis. Specific methods of patent analysis include a method based on descriptive statistics such as frequency analysis, citation analysis and visualization and a method based on inductive statistics such as time series analysis, regression analysis and Bayesian network model. Studies are recently being conducted on the processes of developing new technologies to develop new products through the application of principal component analysis and text mining techniques to keyword data included in patent documents¹².

As today's development of science and technology including IT has brought about a change in the economic paradigm such as the shift of social and economic business areas from the industrial economy to the knowledge-based economy, new value creation through continuous innovation based on technology predictions, the strengthening of core competence, the development of core technology and technological convergence is becoming more important¹³. In a move to adapt soccer-related areas to this current trend, this study intends to identify the trends in soccer-related patents and to make predictions about soccer-related technologies. To this end, this study intends to conduct association rule mining and social network analysis using statistical methods such as principal component analysis and cluster analysis and text mining techniques, based on collected documents. This study will identify the concrete information about core technology factors included in each patent in soccer-related industries and also identify the relationships among technologies and central technologies and patents. Accordingly, it is thought that the findings of this study can be used as basic data for technology predictions.

2. The Current Status of Domestic Sports Industries

In 2015, the Ministry of Culture, Sports and Tourism (MCST) conducted an investigation into a total of 13 items including business information about and business performance (i.e. turnover) of 2,000 businesses sampled with respect to 4 big sports such as soccer, baseball, basketball and volleyball among domestic sports-related businesses. As a result, the Fiscal Year 2014 total turnover

amounted to 4 trillion and 38 billion won and soccer took up the largest portion of the turnover. The turnover of pro sports-related businesses in the 4 big sports amounted to one trillion and 453 billion won, accounting for 36.1% of the total turnover and soccer took up the second largest portion of the turnover after baseball Table 1.

Table 1. Survey of the current status of sports industries

Type	Number of Businesses	Turnover (100 million won)	Number of Workers (persons)
Soccer	4,916	19,970	18,684
Baseball	2,381	10,630	7,571
Basketball	2,586	6,550	10,011
Volleyball	1,751	3,230	3,587
Total	11,634	40,380	39,853

According to the results of a survey of 2,186 sports participants and 3,067 spectators (adults aged 15 years and older) in the 4 big sports regarding the current status of participation and consumption, in soccer, the average period of participation was longest with 8.1 years, related consumption expenditures were the third largest with 1,146,474 won and annual average frequency of watching games was lowest at 4.7 times Table 2.

Table 2. Survey of participant and spectator

Classification	Soccer	Baseball	Basketball	Volleyball
Average period of participation (year)	8.1	5.8	7.8	5.4
annual average frequency of watching game (times)	4.7	7.1	7.7	6.6

Nevertheless, as Korea held the 2002 FIFA World Cup and has advanced to the World Cup finals eight consecutive times, the Korean people's interest in soccer has been steadily growing and soccer-related industries have been continuously developing.

3. Research Methods

3.1 Data Collection

The analysis object data used by this study targeted domestic patents and utilities as 'soccer'-related patents. The extraction of analysis-targeted patents was carried out

using the website (<http://kipris.or.kr>) of Korea Intellectual Property Information Service (hereinafter referred to as KIPRIS) and the period for the search was not subjected to any specific restriction. All patents registered, disclosed and expired were targeted in the process of collection and 185 patents ranging from year 1973 to 2015 based on the year of patent application were selected as analysis objects. KIPRIS used by this study for data collection is a domestic patent search website that enables free searching of information through the data basing of the information on intellectual property rights such as trademark, utility, design and patent by Korea Institute of Patent Information.

3.2 Research Problems

This study intends to investigate the following concrete problems using the patents selected as analysis objects:

First is to create a patent-term data matrix of preceding patents used by each patent in accordance with International Patent Classification (hereinafter referred to as IPC) and based on this, to review the patent trend through the implementation of principal component analysis and cluster analysis.

Second is to identify the patents having a high degree of support, confidence and lift through the implementation of technology association analysis using association rule mining and to carry out technology predictions.

Third is to identify the patents with a high centrality and to discuss the technologies available to future research and development plans, by implementing principal component analysis and fuzzy clustering and based on this, conducting social network analysis.

3.3 Reviews of Relation Theory

As it turned out that it is desirable to use silhouette measures in selecting the optimal number of clusters¹⁴, this study intends to use silhouette measures as the criteria for selecting the optimal number of clusters. The measure of silhouette width is defined as a difference standardized between a (i) and b (i) in an index $s(i) \in [-1, 1]$ against an individual i in a cluster¹⁵. Wherein a (i) indicates an average dissimilarity between an individual i and all the other individuals in a cluster including i. b (i) indicates an average dissimilarity among all individuals in a cluster closest to an individual i. It is decided that as the value of s (i) comes closer to 1, an individual i has a higher similarity to the cluster including i compared with

a neighboring cluster. After calculating the values of s (i) of all individuals numbered n, the optimal number of clusters is selected based on the result of a cluster having the largest $S = \sum s(i)$.

The degree of support, confidence and lift indicating the association rules among IPCs of preceding patents used for analysis-targeted patents is respectively as follows:

$$\text{Support } (X \rightarrow Y) = P(X \cap Y) \\ = (\text{number of transaction including } X \text{ and } Y) / (\text{number of total transaction}),$$

$$0 < \text{Support } (X \rightarrow Y) \leq 1$$

$$\text{Confidence } (X \rightarrow Y) = P(Y | X) = P(X \cap Y) / P(X)$$

$$= (\text{number of transaction including } X \text{ and } Y) / (\text{number of transaction including } X),$$

$$0 \leq \text{Confidence } (X \rightarrow Y)$$

$$\text{Lift } (X \rightarrow Y) = \text{confidence } (X \rightarrow Y) / \text{support } (X \rightarrow Y), \quad 0 \leq \text{lift } (X \rightarrow Y) \leq \infty$$

Generally, social network analysis can identify a central technology which plays a key role in the relevant technology area or typifies this area. A central technology can be effectively used for future research and development plans of states and companies. C (i), the centrality of “closeness” and C_b (γ), the centrality of “betweenness” used by this study are as follows:

$$C(i) = \sum 1/d(i, j) / (n-1), \quad i=1, 2, \dots, n$$

Provided that, since d(i, j) indicates the distance between node i and j, or non-closeness, 1/d(i, j) indicates the closeness between node i and j.

$$C_b(\gamma) = \sum \sum g(ij) / g(ij), \quad \gamma = 1, 2, \dots, n$$

Provided that g (ij) indicates the shortest path between node i and j (≠j) and g (ij) indicates the number of a path passing through node γ (≠i) among the shortest paths between node i and j (≠i, ≠j).

3.4 Analysis Tool

Excel 2010 was used in the whole process of treatment for analysis such as the arrangement of collected patent data and R 3.1.2 was used in the processes of statistics analysis and data mining.

4. Results of Patent Analysis

4.1 Distribution of Patent Application by Years

Table 3 shows the results of classification and arrangement of 185 patents on a 5 year basis to identify the characteristics according to the years of application of collected soccer-related patents. The volume of applications for soccer-related patents has been continuously increasing since a soccer-related patent was first applied for in 1973 and applications for soccer-related patents in Korea was most frequently around 2002 when The 2002 FIFA World Cup was held in Korea. Since then, the number of applications has been decreasing.

Table 3. Distribution of patent application by years

Period	Number of Application	Ratio(%)
1973 to 1990	8	4.3
1991 to 1995	11	5.9
1996 to 2000	38	20.5
2001 to 2005	83	44.9
2006 to 2010	20	10.8
2011 to 2015	25	13.5

4.2 Cluster Analysis

Principal component analysis was conducted targeting 164 patents titled in English among 185 patents collected for soccer-related patents. The principal components accounting for over 90% of the total variations in extracted principal components were selected as the objects of cluster analysis. The number of clusters was finally decided as 3 through the calculation of silhouette measures according to the changes in the number of clusters to select the optimal number of clusters. Subsequently, an analysis of K-medoids clustering under the condition of 3 clusters was conducted for the clustering of 164 patents. As a result, 52 patents (31.7%), 85 patents (51.8%) and 27 patents (16.5%) were allocated and classified into a first cluster, a second cluster and a third cluster respectively. The parent key-words extracted from the patents belonging to each cluster and the central technologies of each cluster utilizing these are defined as shown in Table 4. Considering many situations related to soccer technologies overall, a soccer shoes-related technology belonging to a third cluster is predicted as a technology that will need technology development in the future. In targeting 164 patents, apparatus, game,

goalpost, machine, medium, method, robot, server and shoes were included in top 10 high frequency words.

Table 4. Technology Clustering

cluster	number of patent	main key words	technology definition	proportion (%)
1	52	apparatus, game, machine, medium, method, providing, recording, server, system	soccer game device	31.7
2	85	apparatus, ball, goalpost, method, practice, program, robot, shooting, system	kick exerciser and goalpost	51.8
3	27	shoes, ball, power, spin, sole, structure, improved	soccer shoes	16.5

4.3 Association Technology Extraction

41 transaction data sets and 55 item data sets were created to implement association rule mining to identify a significant relationship among items, targeting patent documents having more than 2 IPC codes among 185 patent documents extracted for patent analysis. The top 10 high frequency IPC codes among 55 IPC codes are shown in Figure 1.

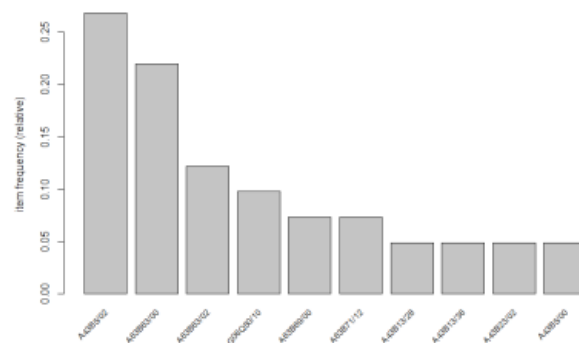


Figure 1. Top 10 high frequency IPC codes.

Next, among the rules with more than 0.001 support degree and more than 0.1 confidence degree, top 10 association rules of the degree of support, confidence and lift are arranged as shown in Tables 5 to 7.

Table 5. Top 10 association rules with high support degree

	lhs	rhs	support	confidence	lift
3	{}	=> {A43B5/02}	0.268293	8 0.2682927	1
2	{}	=> {A63B63/00}	0.219512	0.219512	1
1	{}	=> {A63B63/02}	0.121951	0.121951	1
100	{A63B63/02}	=> {A63B63/00}	0.073171	0.6	2.733333
101	{A63B63/00}	=> {A63B63/02}	0.073171	0.333333	2.733333
81	{A43B13/36}	=> {A43B5/02}	0.04878	1	3.727273
82	{A43B5/02}	=> {A43B13/36}	0.04878	0.181818	3.727273
83	{G06Q50/30}	=> {G06Q50/10}	0.04878	1	10.25
84	{G06Q50/10}	=> {G06Q50/30}	0.04878	0.5	10.25

Table 6. Top 10 Association rules with high confidence degree

	lhs	rhs	support	confidence	lift
4	{A43B21/02}	=> {A43B5/02}	0.02439	1	3.727273
5	{A63H29/22}	=> {A63H17/39}	0.02439	1	41
6	{A63H17/39}	=> {A63H29/22}	0.02439	1	41
7	{A63B71/00}	=> {A63B71/06}	0.02439	1	41
8	{A63B71/06}	=> {A63B71/00}	0.02439	1	41
9	{A63B61/04}	=> {A63B63/00}	0.02439	1	4.555556
11	{G06F17/18}	=> {G06F19/00}	0.02439	1	41
12	{G06F19/00}	=> {G06F17/18}	0.02439	1	41
13	{A61B5/04}	=> {A63B71/12}	0.02439	1	13.66667
15	{D04H13/00}	=> {F24F3/08}	0.02439	1	41

Table 7. Top 10 association rules with high lift degree

	lhs	rhs	support	confidence	lift
5	{A63H29/22}	=> {A63H17/39}	0.02439	1	41
6	{A63H17/39}	=> {A63H29/22}	0.02439	1	41
7	{A63B71/00}	=> {A63B71/06}	0.02439	1	41
8	{A63B71/06}	=> {A63B71/00}	0.02439	1	41
11	{G06F17/18}	=> {G06F19/00}	0.02439	1	41
12	{G06F19/00}	=> {G06F17/18}	0.02439	1	41
15	{D04H13/00}	=> {F24F3/08}	0.02439	1	41
16	{F24F3/08}	=> {D04H13/00}	0.02439	1	41
17	{A43B3/16}	=> {A43B5/18}	0.02439	1	41
18	{A43B5/18}	=> {A43B3/16}	0.02439	1	41

In Tables 5 and 6, a technology with high confidence degree despite low support degree has the likelihood of being vacant technology. The reason is that, although technologies are not yet widely developed, the association between technologies is very high, which means high confidence degree. For example, A43B5/00 is related to 'sports shoes such as slip prevention devices including ice eisers and soccer shoe cleats' and A63H17/39 is related to 'remote control items'. This suggests that the technologies of soccer shoes and online soccer games have the likelihood of being vacant technology.

In Table 7, lift degree more than 1 indicates that

two technologies have mutually complementary relationships with a possible synergistic effect. So, this relationship is open to technology predictions. For example, A63H29/22, which is to provide a driving force through a track using electrical driving devices and A63H17/39, which is a remote control item are found to have mutually complementary relationships in technologies. These technologies are identified as soccer-related robot technologies. Other technologies based on IPC codes can be identified according to the version 2015 of International Patent Classification provided by Korean Intellectual Property Office.

4.4 Technology Network Analysis using Network

Social network analysis is a visualization analysis method to identify in what structures individual nodes are connected internally when they form a network^{16,17}. To identify the network of detailed technologies, it is necessary to analyze the patent for a bipartite network composed of technology clusters¹⁸. As reviewed above, the optimal number of clusters is 3. Fuzzy clustering was implemented after conducting a principal component analysis and finding the scores of principle components. In this case, the patents with the value of a membership, which means that each patent will belong to a certain cluster, more than 0.2 are allocated to 1 and the patents with the value of 0.2 or less are allocated to 0¹⁹. This was used as an affiliation matrix to create a bipartite network.

As shown in Table 8, in the centrality of clusters, all centrality indicators of a second cluster were high. The networks of relationships between the analysis-targeted patents and clusters are shown in Figure 2.

Table 8. Centrality of 3 clusters

centrality	cluster1	cluster2	cluster3
degree	75	107	65
closeness	0.48	0.59	0.45
betweenness	10134	17718	8424

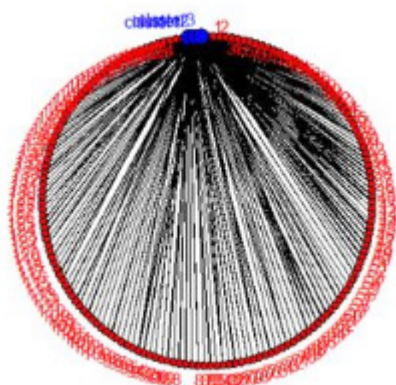


Figure 2. Relationship network between patent and cluster.

The patents belonging to a second cluster mainly include robots or soccer simulators. A network analysis shows that these technologies are relatively significant technologies.

5. Conclusion

Patent information is one of the sources of core technologies most important for research and development. Since patent information includes the commercially available and state-of-the-art technologies, it is the most useful tool in understanding industrial trends and becomes an important reference in establishing the directions of research and development²⁰. As patents are, recently, becoming more significant, the accessibility to patent data also is increasing. Furthermore, the utility of patent analysis also is highly evaluated. At this point, technology predictions were made based on the implementation of an analysis of domestic soccer-related patents through this study. In this regard, this study conducted association rule mining and social network analysis using statistical methods such as principal component analysis and cluster analysis and text mining techniques, based on collected documents. This study result showed the following. First, the result of a cluster analysis showed that apparatus, game, goalpost, machine, medium, method, robot, server, shoes were included in high frequency words. Second, the result of an association analysis showed that the technologies of soccer shoes and online soccer games had the likelihood of being vacant technology. Third, a social network analysis showed that the technologies of robots and soccer simulators were relatively significant technologies. Considering the above results overall, in soccer-related technologies, the technologies related to soccer shoes or robots are predicted as becoming significant technologies in the realm of off-line services and soccer game-related technologies are predicted as becoming significant technologies in the realm of on-line services.

This study identified the relationships among the technologies of soccer-related industries and central technology clusters and patents. It is thought that the findings of this study can be used as basic data to make predictions about the technologies of soccer-related industries.

6. References

1. Jeong WS, Ryum JS, Yi JH. The influence of soccer shoe upper materials on the ball velocity and its torque. The Korean Journal of Physical Education. 2011; 50(6):419–27.

2. Horne VS, Stefanyshyn DJ. Potential method of optimizing the klapskate hinge position in speed skating. *Journal of Applied Biomechanics*. 2005 Aug; 21(3):211–22.
3. Lee EJ, Rim MH. An analysis of foreign sports industry as a promising technology. *The Korea Contents Association Spring Conference*; 2015. p. 367–8.
4. Cha MH. A study for students throw game education contents. *Journal of the Korean Society for Computer Game*. 2011 Oct; 24(4):193–200.
5. Choi J, Kim H, Im N. Keyword network analysis for technology forecasting. *Journal of Intelligence and Information System*. 2011; 17(4):227–40.
6. Weaver WT. The delphi forecasting method. *The Phi Delta Kappan*. 1971 Jan; 52(5):267–71.
7. Bengisu M, Nekhili R. Forecasting emerging technologies with the aid of science and technology databases. *Technological Forecasting and Social Change*. 2006 Sep; 73(7):835–44.
8. Agami NME, Omran AMA, Saleh MM, Shishiny HEEE. An enhanced approach for trend impact analysis. *Technology Forecasting and Social Change*. 2008 Nov; 75(9):1439–50.
9. Lee SJ, Lee SH, Seol HJ, Park YT. Using patent information for designing new product and technology: Keyword based technology roadmapping. *R&D Management*. 2008 Mar; 38(2):169–88.
10. Jun S, Lee SJ. Emerging technology forecasting using new patent information analysis. *International Journal of Software Engineering and its Applications*. 2012 Jul; 6(3):107–16.
11. Choi J, Hwang YS. Patent keyword network analysis for improving technology development efficiency. *Technological Forecasting and Social Change*. 2014 Mar; 83:170–82.
12. Lee S, Yoon B, Park Y. An approach to discovering new technology opportunities: Keywords-based patent map approach. *Technovation*. 2009 Jun-Jul; 29(6-7):481–97.
13. Kho JC. A study on research trend in management of technology using keywords network analysis. [PhD thesis]. Seoul: Sungkyunkwan University; 2013.
14. Jun S, Uhm D. Patent and statistics, what's the connection? *Communications of the Korea Statistical Society*. 2010; 17(2):205–22.
15. Everitt BS, Landau S, Leese M, Stahl D. *Cluster analysis*. 4th ed. New Jersey: Wiley; 2001.
16. Scott JG. *Social network analysis*. 3rd ed. Los Angeles: SAGE; 2012.
17. Huh MH. *Introduction to social network analysis using R*. Seoul: Freedom Academy Press; 2012.
18. Jun S. Technology network model using bipartite social network analysis. *Communications in Computer and Information Science*. 2012 Dec; 340:28–35.
19. Jun S, Park S, Jang D. *Patent analysis at technology forecasting*. Seoul: Kyowoosa; 2014.
20. Kang HJ. A study on the projection of the promising fusion technologies by US patent analysis. [PhD thesis] Seoul: Kookmin University; 2006.