CONVERGENCE IN INFORMATION AND COMMUNICATION TECHNOLOGY (ICT) USING PATENT ANALYSIS

Eunhee Kim
Jaejon Kim
Joon Koh
Chonnam National University, Buk-gu, Gwangju, Korea

_____

ABSTRACT

Since the 1990s, information and communication technology (ICT) has been perceived as the critical technology for economic development, and the ICT industry itself has been growing exceptionally fast. Moreover, technology convergence in ICT has received particular attention. ICT innovations diffuse into existing products and thus come to form a new integral part of the goods. This is an exploratory research to examine technology convergence of the supply side as a firm level in the ICT sector using International Patent Classification (IPC) of 43,636 sample patents from 1995 to 2008. This study finds a degree of merger and relationships between different technology domains through the association rule mining of patent co-classification. This type of analysis helps companies to take strategies under the environment of technological trajectory change.

Keyword: technological convergence; information and communication technology (ICT); patent; international patent classification (IPC); co-classification; association rule mining

_____

Address for correspondence / Endereço para correspondência

Eunhee Kim, is an assistant professor in the Department of Business Administration, Chonnam National University in Korea. She holds a Ph.D. in Technology Management from Seoul National University, a MS in Technology Management from State University of New York at Stony Brook, and a BS in Computer Science from Ewha Womans University. She worked at LG-EDS and EDS Korea as a system engineer and at Industrial Bank of Korea's Economic Research Institute as a research fellow. E-mail: eheekim@chonnam.ac.kr

Jaejon Kim, is a professor in the Department of Business Administration, Chonnam National University. He received his Ph. D. at the Arizona State University. His research interests are in the DSS, GDSS, and e-business areas. E-mail: jaejonkim@hotmail.com

Joon Koh, is an associate professor in the Department of Business Administration, Chonnam National University. He received his Ph.D. in MIS from the Korea Advanced Institute of Science and Technology (KAIST). His research interests are in the areas of virtual community, communities-of-practice, and knowledge management. Chonnam National University, 300 Yongbong-dong, Buk-gu, Gwangju, Korea. E-mail: kjoon@chonnam.ac.kr

1. INTRODUCTION

Since the 1990s, information and communications technology (ICT) has been perceived as the essential technology for economic development, and ICT industries themselves have been growing exceptionally fast. Moreover, technology convergence has received particular attention. ICT innovations diffuse into existing products and thus come to form a new integral part of the goods.

The creation of synergies, blurring of industry boundaries, integration, and overlapping of markets are all used to describe convergence. The convergence phenomenon has been mainly observed and discussed in ICT sectors. ICT innovations diffuse into existing products and thus come to form a new integral part of the goods. Patents play an increasingly important role in innovation and patent data are used to indicate innovative activity of companies, industries and countries. Patent analysis can be regarded as one of the most effective methods to keep in touch with technology trends (Karvonen and Kassi, 2010). Scholars used patent analysis to discuss convergence phenomenon (Duysters and Hagedoorn, 1998; Gambardella and Torrisi, 1998; Bröring, 2005; Curran and Leker, 2011; Kavon and Kassi, 2010).

This is an exploratory research to examine technology convergence of the supply side as a firm level in the ICT sector using patent analysis. International Patent Classification (IPC) of 46,363 sample patents from 1995 to 2008 is employed to find a degree of merger and relationships between different technology fields through the association rule mining of patent co-classification.

The paper is structured as follows: Section 2 outlines shortly the background of convergence and presents the use of patent analysis. Section 3 presents the research setting, including research data and methodology. Section 4 indicates the empirical results. Finally, a discussion of this study and future research strands conclude this paper in Section 5.

2. BACKGROUND

Convergence is an often used but rarely defined concept. Ideas such as the creation of synergies, disappearance of industry boundaries, integration, or overlapping of markets, are all used to describe this phenomenon. Technological convergence is the tendency for different technological systems to evolve towards performing similar tasks (Wikipedia). Convergence can refer to previously separate technologies such as voices, data and videos that now share resources and interact with each other, synergistically creating new efficiencies. The phenomenon of convergence occurs when innovations emerge at the intersection of established and clearly defined industry boundaries, thereby sparking off an evolutionary development with a much broader impact. In recent industry developments within information technology (IT), bio-technology (BT) and nano-technology (NT), the convergence of technologies and knowledge bases has induced a variety of industrial points of inflection. Hence, industry boundaries have become blurred, and innovation does not take place within previously existing industrial silos anymore, but rather between them (Hacklin et al., 2009).

The dictionary definition of convergence is 'tendency to meet at a point' or 'gradual change so as to become similar or develop something in common'. The first use of the term convergence can be traced back to Rosenberg (1963) who introduced the label

'technological convergence' as a way to describe the evolution towards a specialized machine tool industry in the US in the late 1800s. Rosenberg's notion of technological convergence appears to have re-emerged in recent decades as way of describing the apparent merger of telecom, data communication, IT, media and entertainment into a giant ICT and multimedia industry (Gambardella and Torrisi, 1998). During the 90s, convergence was mainly discussed in the context of the merger of the IT, telecommunications, media and entertainment industries into a giant 'infocom' sector (Lind, 2004).

According to a study (Greenstein and Khanna 1997), it is possible to define two basic forms of convergence as substitutes and complements. Competitive convergence as a 'Substitutes Paradigm' occurs when products or services become interchangeable one for another to fulfill a set of certain user needs through bundling of functions. Complementary convergence as a 'Cooperative Paradigm' occurs when products or services from different industries are merged to meet a larger or new set of consumer needs simultaneously. Pennings and Puranam (2001) classified industry convergence from two dimensions such as substitution/complementation and supply/demand. Stieglitz (2003) suggested a similar classification but the dimension of supply/demand with the dimension of technology–based/product-based. Studies of Hacklin (2008) and Hacklin et al (2009) developed and discussed a process of four sequential convergence stages, which are knowledge, technology, application and industry convergence, with an evolutionary perspective. Curran and Leker (2011) discussed how to measure convergence under the sequential process which is evolving when scientific disciplines and technologies and/or markets have converged. Starting with scientific disciplines that begin to use more and more research results of one another, a scientific convergence will start with cross-disciplinary citations and eventually develop further into closer research collaborations. After the distance between basic science areas has been decreasing for some time, applied science and technology development should follow leading to technology convergence (Meyer, 2000; Murray, 2002; Bainbridge, 2006).

Pennings and Puranam (2001) argued that based on a validity assumption for classification schemes like the IPC (International Patent Classification), convergence can be found in patent data through growing overlapping among IPCs and through an increase in patent citations between different classes. Patent analysis has been employed in the context of technology-driven convergence of electronics, computers, and telecommunication (Duysters and Hagedoorn, 1998; Gambardella and Torrisi, 1998; Bröring, 2005) as patents are often regarded as outcome indicators for organizations' R&D activities (Ernst, 1995; Fai and Tunzelmann, 2001). Curran and Leker (2009; 2011) discussed convergence indicators using patent data. Also Xing and et al (2011) tried to measure industry convergence with input-out analysis.

3. RESEARCH METHODS AND SAMPLES

In this research, association rule mining is used to analyze technology integration and diversification in a firm level. Data mining, which is referred to as a knowledge discovery in a database, is a process of nontrivial extraction of implicit previously unknown and potentially useful information such as knowledge rules, constraints and regularities from data in a database (Chen et al., 1996). Data mining, which differs from traditional statistics in that formal statistical inference, is assumption-driven in the sense that a hypothesis is formed and validated against the data. Data mining, in contrast, is discovery-driven in the sense that patterns and hypotheses are automatically extracted from data (Zhang and Zhang, 2002). Data mining has made broad and significant

progress since its early beginning in the 1980's. Today, data mining is used in a vast array of areas, and numerous commercial mining systems are available (Han et al. 2006). Association mining is one of the best-studied methods in data mining (Agrawal et al., 1993; Agrawal and Srikant, 1994; Chen et al., 1996; Han and Kamber, 2001). Since its introduction in 1993 (Agrawal et al.), the area of association rule mining has received a great deal of attention.

Association rule mining has been developed mainly to identify the relationships strongly associated among item sets that have high frequency and strong correlation. Association rules are produced by finding the interesting associations or correlation relationships among a large set of data items (Jiao and Zhang, 2005), and enable us to detect the items that frequently occur together in an application (Zhang and Zhang, 2002).

An association rule (Agrawal et al., 1993) is a probabilistic relationship, of the form A -> B, where A, B are disjoint itemsets. The intuitive meaning of such a rule is that the transactions (or tuples) that contain itemset A also tend to contain itemset B. An association rule indicates that the occurrence of a certain itemset in a transaction will imply the occurrence of another itemset in the same transaction. The rule suggests that a strong relationship exists between the itemsets. The association analysis is applicable to market basket data, bioinformatics, medical diagnosis, Web mining, and scientific data analysis (Tan et al., 2005).

The importance of a rule is usually measured by two numbers, support and confidence. These two properties provide the empirical basis for derivation of the inference expressed in the rule and a measure of the interest in the rule. The support for the association rule A -> B is the percentage of occurrences that contain both itemset A and B among all transactions. The confidence for the rule A -> B is the percentage of transactions that contain an itemset B among the transactions that contain an itemset A (Tan et al., 2005). The rule A -> B holds with support s if s% of transactions in the database contain both itemset A and itemset B. The rule A -> B holds with confidence c if c% of transactions that contain itemset A also contain itemset B. Association rule mining finds all rules in the database that satisfy some minimum support and minimum confidence constraints (Agrawal and Srikant 1994). Additionally, lift value (Brin et al., 1997) is used to judge the strength of an association rule. The lift of a rule A -> B is defined as support (AUB) / (support (A)*support (B)). A lift ratio greater than 1.0 suggests that there is some usefulness to the rule. The larger the lift ratio, the greater the strength of the association (Kim and Park, 2006).

In this study an itemset is regarded as a set which includes primary patent classification code and secondary patent classification codes of a patent, and a transaction means granting a patent. A patent has its classification code, which indicates its technology area. Patent information includes one primary classification code and additionally second classification codes. The patent co-classification shows clearly convergence between different technological domains (Curran and Leker, 2011). In this research, the association rule mining is applied to discover the linkage patterns, which show strongly related among various technology areas, based on the patent co-classification information.

In order to examine the technological convergence, 46,363 granted patents by IBM from 1995 to 2008 in USPTO are used in this study. IBM holds more patents than any other technology company, and has topped the list of the world's most inventive companies from 1993.

The International Patent Classification (IPC) provides for a hierarchical system of language independent symbols for the classification of patents and utility models according to the different areas of technology to which they pertain (WIPO). Each patent grant is assigned to IPC to determine the nature of the patent. One patent can be assigned to more than one IPC if the patent finds application in various domains. Each company has a few subclasses to which most of their patents are assigned. These subclasses describe their core technological competencies. If a company has granted patents only in a few subclasses, it can be said that the technologies employed by the company are highly focused on a narrow field of expertise. On the other hand, if all the patents are not concentrated in a few subclasses, research can be said to be diversified.

The IPC codes in a firm's patent records are identified and classified into technology fields representing the firm's technology domains. Patent application in each field indicates an accumulation of knowledge and advancement in the technological trajectory (Fai and Tunzelmann, 2001). IPC codes are a hierarchical way of assigning the category to which every patent belongs. There are 8 sections, 120 classes, 628 subclasses and about 70,000 groups. The 628 subclasses are aggregated into 35 technological fields, and for descriptive purposes these are further aggregated into five main categories: electrical engineering, instruments, chemistry, mechanical engineering and others, and 35 sub categories in the appendix, the IPC and Technology Concordance Table (WIPO, 2008). The subclasses of the sample patents are analyzed using association rule mining in this research. Among the 43,636 sample patents used for this analysis, 13,338 patents were assigned to more than one IPC..

4. ANALYSIS

In this study, a software package, R, is used to analyze the patent dataset. For the association rule mining, the 'apriori( )' algorithm, which is well known and included in the R package, 'arules' is executed. Thresholds for mining this dataset are 0.05% for minimum support and 90% for minimum confidence. The lift values of all association rules as the results of the association rule mining are greater than 1.0, indicating the usefulness of the rules and the strength of the association.

The table 1 shows the IPC statistics of the sample patent dataset. During fourteen years from 1995 to 2008, IBM's patents were diversified within 355 different technology fields. The major IPC subclass codes of the dataset are 'G06F' and 'H01L' as indicated in table 1. More than 60% of the company's granted patents during this period are included in these two subclasses.

In this study, the conditions for association rule mining of IPC co-classification, minimum support is 0.05% which is lower than a general threshold value. Support value to detect association rules is related to the frequency of occurrence. The main purpose of this study is to examine the technological convergence in a patent using its IPC co-classification. As displayed in the table1, more than half of the company's patents are focused on two technological domains. And others are dispersed into more than 300 different subclasses. In order to detect the association rules between different technology subclasses, the minimum support value should be lowered in this study.

Table 2 shows the results of association rule mining for co-classification of the IPC dataset. The association rules detected from the patent dataset shows clearly the relationships between different fields of technology. As mentioned above, subclasses of IPC can be grouped into 35 technology fields as IPC and Technology concordance table (WIPO, 2008). The 32 association rules can be interpreted based on the concordance

table. The detected 32 association rules can be divided into five types. First, the association rules from #1 to #17 present the relationships between 'computer technology' and other fields such as 'telecommunications', 'digital communication' and 'semiconductors' in electrical engineering. Second, the association rules from #18 to #26 show the technological combinations between electrical engineering and instrument. Third, the association rule #27 indicates the technological integration among three domains: electrical engineering, instruments, and mechanical engineering fields. Fourth, the association rules #28, #29, and #30 show the technological combinations between electrical engineering and mechanical engineering. Fifth, the last two association rules #31 and #32 present the relationships between 'organic fine chemistry' and 'biotechnology' in chemistry. These two rules do not designate the convergence between 'computer technology' which the company are focused on and technology fields in chemistry. They indicate the combinations between 'organic fine chemistry' and 'macromolecular chemistry, polymers' fields.

Table 1. Patent Statistics for major IPC subclasses of sample dataset (n=43,636)

Year	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	Total	%
Count of patent	155	364	1,666	3,495	4,411	4,432	4,067	3,599	3,996	3,695	3,602	3,451	4,443	4,987	46,363	100.0%
Major IPC subclass																
G06F	77	183	876	1,720	2,093	2,087	1,778	1,781	2,070	1,946	1,872	1,763	2,289	2,647	23,182	50.0%
H01L	12	20	97	288	438	452	455	386	548	542	543	538	671	626	5,616	12.1%
H04L	1	13	108	113	157	155	165	131	153	146	135	125	144	174	1,720	3.7%
G11B	18	26	96	206	184	185	153	135	78	55	77	49	70	92	1,424	3.0%
G01R	1	8	29	72	81	65	69	56	83	112	102	105	125	137	1,045	2.2%
G11C		4	11	44	61	79	101	73	106	76	103	86	163	133	1,040	2.2%
H05K	1	8	24	74	86	69	79	70	54	61	54	63	71	107	821	1.7%
G06K	5	7	33	60	86	84	58	54	42	50	53	52	61	87	732	1.5%
H03K	1	5	9	52	67	61	65	38	56	56	45	51	53	59	618	1.3%
G09G	4	14	28	65	105	103	96	55	22	13	10	10		15	553	1.1%
G06Q				5	8	52	81	55	48	46	40	52	57	65	509	1.1%
G10L		4	12	52	86	82	75	41	49	27	17	13	23	19	500	1.0%
H04M	1	2	14	22	21	42	66	58	57	53	34	34	25	34	463	1.0%

Table 2. Association rules* for IPC co-classification of sample dataset**

No.	Antecedent		Consequent	Support	Confidence	Technology categories***
1	{G06F,G06K,H04K,H04N}	=>	{H04L}	0.0005	1.0000	I
2	{G06F,G06K,H04L,H04N}	=>	{H04K}	0.0005	1.0000	
3	{G06K,H04K,H04L,H04N}	=>	{G06F}	0.0005	1.0000	
4	{G06K,H04K,H04N}	=>	{H04L}	0.0005	1.0000	

5	{G06K,H04K,H04N}	=>	{G06F}	0.0005	1.0000	
6	{G06K,H04L,H04N}	=>	{H04K}	0.0005	1.0000	
7	{G06K,H04L,H04N}	=>	{G06F}	0.0005	1.0000	
8	{G06F,H04K,H04N}	=>	{H04L}	0.0007	1.0000	
9	{H04K,H04L,H04N}	=>	{G06F}	0.0007	1.0000	
10	{G06F,G06K,H04K}	=>	{H04L}	0.0010	1.0000	
11	{G06F,G08C,H04L}	=>	{H04J}	0.0016	0.9565	
12	{G06K,H04K}	=>	{H04L}	0.0010	0.9333	
13	{G06K,H04K,H04L}	=>	{G06F}	0.0010	0.9286	
14	{G06N}	=>	{G06F}	0.0092	0.9248	
15	{G08C,H04J,H04L}	=>	{G06F}	0.0016	0.9167	
16	{H04K,H04N}	=>	{H04L}	0.0007	0.9000	
17	{H04K,H04N}	=>	{G06F}	0.0007	0.9000	
18	{G01R,G06F,G08C,H04L}	=>	{H04J}	0.0014	1.0000	I, II
19	{G01R,G08C,H04J,H04L}	=>	{G06F}	0.0014	1.0000	
20	{G01R,G08C,H04L}	=>	{H04J}	0.0014	1.0000	
21	{G01R,G08C,H04L}	=>	{G06F}	0.0014	1.0000	
22	{G01R,G08C,H04J}	=>	{G06F}	0.0016	1.0000	
23	{G01R,G08C}	=>	{G06F}	0.0017	1.0000	
24	{G01R,G06F,G08C}	=>	{H04J}	0.0016	0.9565	
25	{G01R,G08C}	=>	{H04J}	0.0016	0.9565	
26	{G08B,H04L}	=>	{G06F}	0.0013	0.9000	
27	{G06F,G21K}	=>	{G03F}	0.0006	1.0000	I, II,IV
28	{G21C}	=>	{G06F}	0.0012	1.0000	I,IV
29	{F28F,H01L}	=>	{H05K}	0.0015	1.0000	
30	{F28F}	=>	{H05K}	0.0043	0.9661	
31	{C07H}	=>	{C12Q}	0.0005	1.0000	III
32	{C12M}	=>	{C12Q}	0.0006	1.0000	

* Thresholds for association rule mining; minimum support - 0.05%, minimum confidence - 90%

** The number of patents with more than one IPC: 13,338
*** 5 technology categories: I. Electrical engineering II. Instruments III. Chemistry IV. Mechanical engineering

5. DISCUSSION AND FUTURE DIRECTIONS

This study aims at examination of technology convergence of the supply side in the ICT sector through patent IPC analysis. Based on IPC of 46,363 sample patents from 1995 to 2008 in ICT, association rule mining of the sample patents' IPC is used to find degree of overlaps and relationships between different technology domains. The results of association rule mining of the ICT firm's patent co-classification show clearly

convergence between different technological domains. Technological convergence can trigger market convergence with new product and firms begin to merge with each other, completing the convergence process with industry fusion, considering sequential process of convergence. Curran and Leker (2011) discussed the phase of convergence based on the assumption of an idealized time series of events: scientific/knowledge convergence, technology convergence, market/applicational convergence, and industry convergence. Also, as scholars suggested in the previous studies, convergence can be considered in several dimensions: supply/demand, substitution/complementation, and product-based/ technology-based.

The context of this study is the ICT technology convergence in a firm level as a technology supplier, considering degree and scope of technology convergence. The results imply that the technology convergence in a firm occurs mainly within its dominant technology areas and that the major technologies tend to be merged with other areas' technology, expending the scope of convergence. However, the results do not indicate the technological convergence change technological paradigm in a firm level. The scope of the convergence can be related to the firm's capability of innovativeness and technological competitiveness. As discussed in previous literature (Curran and Leker, 2011), convergence starts with knowledge convergence. Additionally, this study examines the usefulness of association rule mining to indicate technological convergence. Due to the technology development, new phenomena have appeared in the world. Some of them can be difficult to be described or analyzed by conventional methods. The association rule mining approach is appropriate for describing the complicate relational data and discovering important patterns among them. The association rule mining analysis can be applied to indicate not only technology convergence but also knowledge convergence.

Technological convergence, as an emerging research field, is being studied by many scholars. The impact of this new phenomenon is enormous. For example, after introducing the smartphone which is a representative outcome of the convergence, lots of new business models and applications related to the smartphone have appeared, and our life style has changed rapidly. In order to improve the capability of forecasting the technology development, in-depth understanding of the technological convergence phenomenon is essential. Even though, in this research, one factor, which is classification code of patent, is considered to discuss technological convergence, knowledge transfer with a perspective of open innovation can be deliberated together using the mining analysis. Therefore, the empirical analysis about the relationship between technology convergence and knowledge will be discussed in a further study.

REFERENCES

Agrawal, R., Imielinski, T., Swami, A. (1993) 'Mining association rules between sets of items in large databases', Proceedings of ACM SIGMOD, 207-216.

Agrawal, R., Srikant, R. (1994) 'Fast algorithms for mining association rules', Proceedings of the 20th international conference on very large databases, 487-499.

Bainbridge, W.S., (2006) 'Transformative Concepts in Scientific Convergence', Annals of the New York Academy of Sciences 1093, 24–45.

Brin, S., Motwany, R., Ullman, J.D. and Tsur, S. (1997) 'Dynamic itemset counting and implication rules for market basket data', Proceedings of ACM SIGMOD, pp.255–264.

Bröring, S. (2005) The Front End of Innovation in Converging Industries: The case of nutraceuticals and functional foods, DUV, Wiesbaden, Germany, 342.

Curran, C., Leker, J. (2009) 'Seeing the Next iPhone Coming Your Way: How to Anticipate Converging Industries', PICMET 2009 Proceedings, August 2-6, Portland, Oregon USA.

Curran, C., Leker, J. (2011) 'Patent indicators for monitoring convergence – examples from NFF and ICT', Technological Forecasting & Social Change 78, 256-273.

Duysters, G., Hagedoorn, J. (1998) 'Technological convergence in the IT industry: the role of strategic technology alliances and technological competencies', International Journal of the Economics of Business 5 (3), 355–368.

Ernst, H. (1995) 'Patenting strategies in the german mechanical engineering industry and their relationship to company performance', Technovation 15 (4), 225–240.

Fai, F., Tunzelmann, V. N. (2001) 'Industry-specific competencies and converging technological systems: evidence from patents', Structural change and Economic Dynamics (12), 141-170.

Gambardella, A., Torrisi, S. (1998) 'Does technological convergence imply convergence in markets? Evidence from the electronics industry', Research Policy 27, 445–463

Greenstein, S., Khanna, T. (1997) 'What does industry convergence mean', in Yoffie, D. (ed.) (1997) Competing in the age of digital convergence, Boston, MA, Harvard Business School Press.

Hacklin, F. (2008) Management of Convergence in Innovation – Strategies and Capabilities for Value Creation Beyond Blurring Industry Boundaries, Physica-Verlag, Heidelberg

Hacklin, F., Marxt, C., Fahrni, F. (2009) 'Coevolutionary cycles of convergence: An extrapolation from the ICT industry', Technological Forecasting & Social Change 76, 723–736.

Han, J., Kamber, M., & Pei, J. (2006) Data mining: concepts and techniques. Morgan kaufmann.

Jiao, J., Zhang, Y. (2005) 'Product portfolio identification based on association rule mining', Computer-Aided Design 37(2), 149-172.

Karvonen, M., Kässi, T. (2010) 'Patent Citations as a Method for Analysing Industrial Convergence', PICMET 2011 Proceedings of Technology Management for Global Economic Growth.

Kim, E. H., Park, Y. (2006) 'An exploratory study of risks in information system development projects: using association rule mining', International Journal of Technology Intelligence and Planning, 2(4), 404-417.

Meyer, M. (2000) 'Does science push technology? Patents citing scientific literature', Research Policy 29 (3), 409–434.

Lind, J. (2004) 'Convergence: History of term usage and lessons for firm strategists', In ITS 15th Biennial Conference, Berlin, Germany, International Telecommunications Society (ITS).

Murray, F. (2002) 'Innovation as co-evolution of scientific and technological networks: exploring tissue engineering', Research Policy 31 (8–9), 1389–1403.

Pennings, J. M., Puranam, P. (2001) 'Market convergence & firm strategy: new directions for theory and research', ECIS Conference, The Future of Innovation Studies, Eindhoven, Netherlands.

Rosenberg, N. (1963) 'Technological Change in the Machine Tool Industry, 1840–1910', The Journal of Economic History 23(4), 414-443.

Stieglitz, N. (2003) 'Digital Dynamics and Types of Industry Convergence – The Evolution of the Handheld Computers Market in the 1990s and beyond', in Christensen F. J., Maskell P. (eds.) The Industrial Dynamics of the New Digital Economy, London: Edward Elgar.

Tan, P., Steinbach, M. Kumar, V. (2005) Introduction to Data Mining, Addison-Wesley.

Wing, W., Ye, X., Kui, L.. (2011) 'Measuring convergence of China's ICT industry: An input-output analysis', Telecommunications Policy 35, 301-313.

WIPO (2008) World Patent Report: A Statistical Review

Zhang, C., Zhang, S. (2002) Association Rule Mining: models and algorithms, Berlin: Springer.

Appendix

IPC code descriptions of sample dataset

IPC code	Description
G	**PHYSICS**
G01	MEASURING; TESTING
G01R	MEASURING ELECTRIC VARIABLES; MEASURING MAGNETIC VARIABLES
G06	COMPUTING; CALCULATING; COUNTING
G06F	ELECTRIC DIGITAL DATA PROCESSING
G06K	RECOGNITION OF DATA; PRESENTATION OF DATA; RECORD CARRIERS; HANDLING RECORD CARRIERS
G06Q	DATA PROCESSING SYSTEMS OR METHODS, SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES; SYSTEMS OR METHODS SPECIALLY ADAPTED FOR ADMINISTRATIVE, COMMERCIAL, FINANCIAL, MANAGERIAL, SUPERVISORY OR FORECASTING PURPOSES, NOT OTHERWISE PROVIDED FOR
G09	EDUCATING; CRYPTOGRAPHY; DISPLAY; ADVERTISING; SEALS
G09G	ARRANGEMENTS OR CIRCUITS FOR CONTROL OF INDICATING DEVICES USING STATIC MEANS TO PRESENT VARIABLE INFORMATION
G10	MUSICAL INSTRUMENTS; ACOUSTICS
G10L	SPEECH ANALYSIS OR SYNTHESIS; SPEECH RECOGNITION; AUDIO ANALYSIS OR PROCESSING
G11	INFORMATION STORAGE
G11B	INFORMATION STORAGE BASED ON RELATIVE MOVEMENT BETWEEN RECORD CARRIER AND TRANSDUCER
G11C	STATIC STORES
H	**ELECTRICITY**
H01	BASIC ELECTRIC ELEMENTS
H01L	SEMICONDUCTOR DEVICES; ELECTRIC SOLID STATE DEVICES
H03	BASIC ELECTRONIC CIRCUITRY
H03K	PULSE TECHNIQUE
H04	ELECTRIC COMMUNICATION TECHNIQUE
H04L	TRANSMISSION OF DIGITAL INFORMATION, e.g. TELEGRAPHIC COMMUNICATION
H04M	TELEPHONIC COMMUNICATION
H05	ELECTRIC TECHNIQUES NOT OTHERWISE PROVIDED FOR
H05K	PRINTED CIRCUITS; CASINGS OR CONSTRUCTIONAL DETAILS OF ELECTRIC APPARATUS; MANUFACTURE OF ASSEMBLAGES OF ELECTRICAL COMPONENTS

IPC and Technology Concordance Table (WIPO, 2008)

Field of Technology	International Patent Classification (IPC) Symbols
I: Electrical engineering	
Electrical machinery, apparatus, energy	F21#, H01B, H01C, H01F, H01G, H01H, H01J, H01K, H01M, H01R, H01T, H02#, H05B, H05C, H05F, H99Z
Audio-visual technology	G09F, G09G, G11B, H04N-003, H04N-005, H04N-009, H04N-013, H04N-015, H04N-017, H04R, H04S, H05K
Telecommunications	G08C, H01P, H01Q, H04B, H04H, H04J, H04K, H04M, H04N-001, H04N-007, H04N-011, H04Q
Digital communication	H04L
Basic communication processes	H03#
Computer technology	(G06# not G06Q), G11C, G10L`
IT methods for management	G06Q
Semiconductors	H01L
II: Instruments	
Optics	G02#, G03B, G03C, G03D, G03F, G03G, G03H, H01S
Measurement	G01B, G01C, G01D, G01F, G01G, G01H, G01J, G01K, G01L, G01M, (G01N not G01N-033), G01P, G01R, G01S; G01V, G01W, G04#, G12B, G99Z
Analysis of biological materials	G01N-033
Control	G05B, G05D, G05F, G07#, G08B, G08G, G09B, G09C, G09D
Medical technology	A61B, A61C, A61D, A61F, A61G, A61H, A61J, A61L, A61M, A61N, H05G
III: Chemistry	
Organic fine chemistry	(C07B, C07C, C07D, C07F, C07H, C07J, C40B) not A61K, A61K-008, A61Q
Biotechnology	(C07G, C07K, C12M, C12N, C12P, C12Q, C12R, C12S) not A61K
Pharmaceuticals	A61K not A61K-008
Macromolecular chemistry, polymers	C08B, C08C, C08F, C08G, C08H, C08K, C08L
Food chemistry	A01H, A21D, A23B, A23C, A23D, A23F, A23G, A23J, A23K, A23L, C12C, C12F, C12G, C12H, C12J, C13D, C13F, C13J, C13K
Basic materials chemistry	A01N, A01P, C05#, C06#, C09B, C09C, C09F, C09G, C09H, C09K, C09D, C09J, C10B, C10C, C10F, C10G, C10H, C10J, C10K, C10L, C10M, C10N, C11B, C11C, C11D, C99Z
Materials, metallurgy	C01#, C03C, C04#, C21#, C22#, B22#
Surface technology, coating	B05C, B05D, B32#, C23#, C25#, C30#
Micro-structural and nano-technology	B81#, B82#
Chemical engineering	B01B, B01D-000#, B01D-01##, B01D-02##, B01D-03##, B01D-041, B01D-043, B01D-057, B01D-059, B01D-06##, B01D-07##, B01F, B01J, B01L, B02C, B03#, B04#, B05B, B06B, B07#, B08#, D06B, D06C, D06L, F25J, F26#, C14C, H05H
Environmental technology	A62D, B01D-045, B01D-046, B01D-047, B01D-049, B01D-050, B01D-051, B01D-052, B01D-053, B09#, B65F, C02#, F01N, F23G, F23J, G01T, E01F-008, A62C
IV: Mechanical engineering	
Handling	B25J, B65B, B65C, B65D, B65G, B65H, B66#, B67#
Machine tools	B21#, B23#, B24#, B26D, B26F, B27#, B30#, B25B, B25C, B25D, B25F, B25G, B25H, B26B
Engines, pumps, turbines	F01B, F01C, F01D, F01K, F01L, F01M, F01P, F02#, F03#, F04#, F23R, G21#, F99Z
Textile and paper machines	A41H, A43D, A46D, C14B, D01#, D02#, D03#, D04B, D04C, D04G, D04H, D05#, D06G, D06H, D06J, D06M, D06P, D06Q, D99Z, B31#, D21#, B41#
Other special machines	A01B, A01C, A01D, A01F, A01G, A01J, A01K, A01L, A01M, A21B, A21C, A22#, A23N, A23P, B02B, C12L, C13C, C13G, C13H, B28#, B29#, C03B, C08J, B99Z, F41#, F42#
Thermal processes and apparatus	F22#, F23B, F23C, F23D, F23H, F23K, F23L, F23M, F23N, F23Q, F24#, F25B, F25C, F27#, F28#
Mechanical elements	F15#, F16#, F17#, G05G
Transport	B60#, B61#, B62#, B63B, B63C, B63G, B63H, B63J, B64#
V: Other fields	
Furniture, games	A47#, A63#
Other consumer goods	A24#, A41B, A41C, A41D, A41F, A41G, A42#, A43B, A43C, A44#, A45#, A46B, A62B, B42#, B43#, D04D, D07#, G10B, G10C, G10D, G10F, G10G, G10H, G10K, B44#, B68#, D06F, D06N, F25D, A99Z
Civil engineering	E02#, E01B, E01C, E01D, E01F-001, E01F-003, E01F-005, E01F-007, E01F-009, E01F-01#, E01H, E03#, E04#, E05#, E06#, E21#, E99Z
Definition for energy technology	International Patent Classification (IPC) Symbols
Solar energy (includes solar photovoltaic power and solar thermal power)	F03G 6/06, F24J 2/00, F24J 2/02, F24J 2/04, F24J 2/05, F24J 2/06, F24J 2/07, F24J 2/08, F24J 2/10, F24J 2/12, F24J 2/13, F24J 2/14, F24J 2/15, F24J 2/16, F24J 2/18, F24J 2/23, F24J 2/24, F24J 2/36, F24J 2/38, F24J 2/42, F24J 2/46, F03G 6/06, G02B 5/10, H01L 31/052, E04D 13/18, H01L 25/00, H01L 31/04, H01L 31/042, H01L 31/052, H01L 31/18, H02N 6/00, E04D 1/30, G02F 1/136, G05F 1/67, H01L 25/00, H01L 31/00, H01L 31/042, H01L 31/048, H01L 33/00, H02J 7/35, H02N 6/00
Fuel cells technology	H01M 4/00, H01M 4/86, H01M 4/88, H01M 4/90, H01M 8/00, H01M 8/02, H01M 8/04, H01M 8/06, H01M 8/08, H01M 8/10, H01M 8/12, H01M 8/14, H01M 8/16, H01M 8/18, H01M 8/20, H01M 8/22, H01M 8/24
Wind energy technology	F03D*, B60L 8/00