

특허분석을 이용한 지능형시스템의 기술예측

Technology Forecasting of Intelligent Systems using Patent Analysis

전성해
Sunghae Jun

청주대학교 바이오정보통계학과

요 약

학습과 추론을 이용하여 주어진 환경에서 최적의 문제해결을 수행하는 지능형시스템에 대한 사용은 날로 증가하고 있다. 다양한 분야에서 지능형시스템은 과거와 현재 뿐만 아니라 미래에도 인간의 삶의 질을 향상시키는 중요한 역할을 수행하기 때문에 이에 대한 기술적 동향과 과학 기술예측은 중요하게 다루어져야 할 것이다. 본 논문에서는 지금까지 개발된 기술에 대한 객관적이고 광범위한 정보를 포함하고 있는 특허데이터를 정량적으로 분석하여 지능형시스템에 대한 기술적 예측에 필요한 특허분석 방법을 제안한다. 지금까지 출원, 등록된 특허데이터를 이용하여 제안방법에 대한 성능평가를 수행한다.

Abstract

Needs of intelligent system has risen continuously to solve given problem optimally using learning and reasoning. This system has performed important roles in diverse fields for improving the human-life quality in past, present, and future. So, it is important to analyze the trend of technology forecasting for the intelligent system. In this paper, we propose a patent analysis method for technology forecasting of the intelligent system using objective patent data. To verify our study, we use the patent data applied and registered until now.

Key Words : 지능형시스템, 특허군집화, 특허분석, 기술예측

1. 서 론

데이터로부터의 학습(learning from data)을 통한 추론(reasoning)에 의해 얻게 된 지능(intelligence)을 시스템에 적용하여 지능형시스템(intelligent system)을 구축할 수 있다. 이와 같은 지능형시스템에 대한 연구는 현재 학교뿐만 아니라 국가 및 민간기업의 연구기관에서 활발히 진행되고 있다. 다양한 분야에서 지능형시스템은 인간의 삶의 질 향상에 기여해 오고 있다. 이 분야에 대한 관심과 연구는 아주 오래전부터 연구자들에 의해 이루어져 왔다. 1940년대에 신경망(neural networks) 모형이 발표되고 이를 바탕으로 한 복잡한 계산이 가능해지면서 지능형시스템에 대한 연구는 더욱 활기를 띠게 되었다[1]. 지금까지 지능형시스템에 대한 기술연구는 양적으로나 질적으로 상당부분 이루어지고 있다. 하지만 지능형시스템을 세부기술 분야별로 살펴보면 활발한 연구개발이 이루어지는 분야와 그렇지 못한 분야가 존재한다. 특히 앞으로 필요할 세부기술에 대한 연구가 제대로 이루어지지 못하면 이에 대한 기술수요에 적절히 대응할 수 없게 된다. 본 논문에서는 과거부터 현재까지 지능형시스템에 대한 연구, 개발 결과를 이용하여 향후 필요한 공백기술을 예측한다. 이를 통하여 지능형시스템에 대한 학교와 기업 그리고 국가의 연구개발 계획에 도움이 될 수 있는 결과를 찾아낸다. 아울러 불필요한 특허소송을 피하고

중복된 기술에 대한 불필요한 연구, 개발을 방지할 수 있게 될 것이다[2],[3]. 본 연구에서 제안하는 방법의 성능을 평가하기 위하여 미국특허청으로부터 지금까지 출원, 등록된 지능형시스템에 관한 전체기술특허를 이용하여 이 분야에 대한 공백기술을 예측한다.

2. 특허데이터와 기술예측

특허는 출원한 기술을 나타낸 도면인 그림과 출원한 기술을 설명한 문자로 이루어진다[4]. 출원번호, 공개번호, 등록번호 등 해당특허를 표시하는 번호와 출원날짜, 공개날짜, 등록날짜 등 시간을 나타내는 숫자가 포함되어 있다. 특허특허문서에 포함된 여러 정보 중에서 해당특허가 인용한 특허와 해당특허를 인용한 특허정보는 현재 기술의 연관성 분석을 위한 중요한 정보로 이용된다. 따라서 대부분의 특허데이터의 정보는 문자와 숫자로 이루어져 있는 문서형태로 존재한다. 특허문서는 포함된 정보의 종류에 따라 해당 특허청으로부터 공식적으로 인정된 정보를 포함한 문서, 독점적 특허정보에 관한 문서, 라이선스와 특허권의 유지에 관한 문서 등으로 구별된다. 이와 같은 문서들을 이용하여 기술예측의 정량적 분석을 하기 위해서는 특허문서를 분석 가능한 데이터형태로 변환시키는 전처리(pre-processing) 작업이 필요하다. 최근에 특허데이터를 이용한 기술예측의 연구결과들이 발표되고 있지만 정량적 데이터분석의 한계 때문에 해당 기술분야에 종사하는 전문가의 주관적인 판단에 의존하고 있다[5],[6],[7],[8],[9],[10],[11],[12],[13]. 본 논문에서

접수일자 : 2010년 8월 21일
완료일자 : 2010년 11월 30일

서는 좀 더 객관적인 특허데이터의 분석을 위하여 다변량통계의 차원축소와 특징추출 그리고 데이터마이닝의 군집화 기법을 이용한다[1]. 이를 통하여 좀 더 객관적인 기술예측 과정을 제안한다. 특히 본 논문에서는 지능형시스템에 관한 공백기술예측을 위하여 지금까지 출원, 등록된 지능형시스템 기술관련 전체특허에 대하여 본 논문에서 제안하는 객관적이고 정량적인 기법을 이용하여 분석한다.

3. 지능형시스템 기술예측을 위한 특허분석

우리나라의 KIPRIS(Korea Intellectual Property Rights Information Service)[14]. 또는 미국의 USPTO(United States Patent and Trademark Office)[15]에 제공하는 특허데이터베이스로부터 검색된 특허문서는 그 자체로는 정량적 분석이 어렵기 때문에 텍스트 마이닝(text mining)과 같은 전처리 절차를 통하여 분석이 가능한 데이터구조로 변환해야 한다[16],[17],[18]. 본 논문에서는 분석이 가능한 데이터구조 중의 하나인 특허-키워드행렬(patent - keyword matrix)을 이용한다. 본 논문에서 사용되는 이와 같은 행렬 데이터는 각 특허문서의 초록을 이용하여 다음과 같은 행렬을 구축한다.

$$X_{n \times p} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix} \quad (1)$$

n 과 p 는 각각 특허수와 키워드수를 나타낸다. x_{ij} 는 i 번째 특허에서 나타난 j 번째 키워드의 발생회수를 의미한다. 일반적으로 특허-키워드행렬에서 특허수보다 키워드수가 큰 경향($n \ll p$)을 보인다. 하지만 행렬을 이용한 데이터 분석이 가능하기 위해서는 행의 크기가 열의 크기보다 훨씬 커야($n \gg p$)한다[19]. 또한 행렬 X 는 매우 희소한(sparse) 데이터구조를 갖기 때문에 이 문제를 해결하기 위하여 본 논문에서는 한 행렬(dimension reduction)를 고려한다. 대표적인 한 행렬기법은 [1]량통계분석 기법인 주성분분석(principal component analysis)이다[19]. 이와 같은 차원축소 과정을 거쳐 다음과 같은 특허-특징행렬(patent - feature matrix)을 구한다.

$$P_{n \times r} = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1r} \\ p_{21} & p_{22} & \dots & p_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nr} \end{pmatrix} \quad (2)$$

특징 p 는 전체키워드에 대한 선형결합의 형태로 표시된다. r 은 특징으로 표현된 주성분의 개수이다. 각 키워드에 대한 가중치는 다음과 같이 벡터 a 로 표시된다.

$$P = \tilde{a}X \quad (3)$$

하지만 특허-키워드행렬에서 열의 수가 행의 수보다 크기 때문에 위의 주성분분석을 위한 행렬연산은 어렵다. 이 문제를 해결하기 위하여 본 논문에서는 SVD(singular value decomposition)을 사용한다. SVD는 행렬데이터의 분석에서 나타내는 계산상의 문제점을 해결하는 기법이다[23]. 분석이 어려운 원래의 데이터구조에 대하여 SVD를 적용하면 쉽게 행렬계산의 문제를 해결할 수 있다. 처음에 SVD는 선형방정식(linear equation)의 최소제곱(least square)문제

해결에서 시작하여 지금은 다양한 계산문제에서 사용되고 있다[20]. 두 개의 벡터 x 와 y 가 직교(orthogonal)하면 다음이 성립한다.

$$x^t y = 0 \quad (4)$$

위 식에서 t 는 벡터와 행렬의 전치(transpose)를 의미한다. A 가 정방행렬(square matrix)이고 모든 열(column)들의 길이가 1인 직교벡터($x^t x = 1$)이면 A 는 직교행렬(orthogonal matrix)이 되어 다음을 만족한다.

$$A^t A = I \quad (5)$$

위 식에서 I 는 단위행렬(identity matrix)이다. 행과 열의 차원이 각각 M 과 N 인 행렬 A 의 SVD는 다음과 같은 분해(factorization)로 정의된다.

$$A = U \Sigma V^t \quad (6)$$

위 식에서 U 는 ($M \times M$)의 직교행렬이고 V 는 ($N \times N$)의 직교행렬이다. 그리고 Σ 는 ($M \times N$)의 대각행렬(diagonal matrix)이고 각각의 원소(element)는 다음과 같다.

$$s_{ij} = 0 \text{ if } i \neq j \text{ and } s_{ii} = s_i \geq 0 \quad (7)$$

또한 $s_1 \geq s_2 \geq \dots \geq s_N \geq 0$ 인 비유일(non-unique)행렬인 U 와 V 가 존재한다. 이 때 s_i 는 A 의 정칙값(singular value)이 되고 U 와 V 의 열들은 각각 좌, 우 정칙벡터(left and right singular vectors)가 된다. 이와 같은 행렬분해를 통하여 계산이 어려운 행렬 A 가 U 와 V 로 분해되어 분석을 위한 행렬계산을 가능하게 한다.

다음으로 본 논문에서 기술예측을 위한 특허데이터의 군집화를 수행하기 위한 최적 군집수 선택의 측도로서 BIC(Bayesian information criterion)를 이용한다. BIC는 모수(parameter)의 변화에 따라 모형의 성능을 평가하여 최적의 모형을 선택할 수 있도록 한다[19]. 일반적으로 최대우도추정(maximum likelihood estimation)에 의한 모수선택은 과대적합(overfitting)의 문제가 발생한다. 하지만 BIC는 모수의 수에 대한 페널티(penalty)를 도입하여 이 문제를 해결하였다. 주어진 모형의 BIC는 다음과 같이 구한다.

$$BIC = -2 \ln L + k \ln(n) \quad (8)$$

위 식에서 x 는 관측된 데이터이고, n 은 x 의 표본크기(sample size)이다. k 와 L 은 각각 추정되어지는 자유모수(free parameters)의 수와 우도함수의 최대값을 나타낸다. 모형에 대한 정규분포(Normal distribution)를 가정하면 위 식은 다음과 같이 n 에 의존하는 식으로 나타낼 수 있다.

$$BIC = \ln(\sigma_c^2) + \frac{k}{n} \ln(n) \quad (9)$$

위 식에서 σ_c^2 은 오차의 분산이다. 주어진 2개의 모형 중에서 더 작은 BIC값을 갖는 모형이 선택된다. 군집분석에서도 BIC값을 최소화하는 군집수에 의해 군집화가 이루어진다. 왜냐하면 군집화는 군집내 분산은 최소로, 군집간 분산은 최대로 이루어지게 실행되기 때문이다. 본 논문에서는 SVD에 의한 차원축소와 BIC에 의한 군집화를 이용하여 지능형시스템에 관한 특허데이터를 분석하고 이를 바탕으로 향후 지능형시스템에 대한 공백기술을 예측한다. 다음은 본 연구에서 제안한 방법에 대한 단계별 과정이다.

- step1) 검색식을 이용한 특허데이터의 획득(한국특허청(KIPRIS), 미국특허청(USPTO), ...)
- step2) 특허문서를 특허-특징행렬(patent-feature matrix)로 변환
- step3) 상위 10개의 주성분을 이용한 특허-특징행렬의 군집화
 - 3-1) BIC 측도를 이용한 최적군집수(optimal number of clusters) K 결정
 - 3-2) 3-1)에서 구한 K값을 이용하여 K-means 군집화 수행
- step4) K개의 군집에 속한 특허들을 이용하여 각 군집의 기술정의
- step5) step4)의 결과를 이용하여 공백기술 선정

4. 실험 및 결과

지능형시스템 기술에 관한 전체특허를 검색하기 위하여 이 분야에 대한 후보 키워드들을 사용하여 최종적으로 다음과 같은 키워드 검색식을 사용하였다.

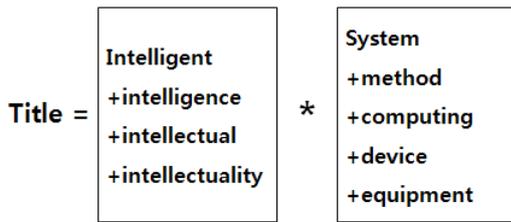


그림 1. 키워드 검색식
Fig. 1. Keywords searching

위 식에서 +는 or 연산자이고 *는 and 연산자이다. 본 논문에서는 위 검색식을 이용하여 가장 많은 유효특허를 검색할 수 있었다. 지능형시스템 분야에 대한 기술예측을 위한 분석데이터를 얻기 위하여 KIPRIS로부터 2010년 5월 17일까지 출원, 등록된 전체 미국특허를 검색하였다[14],[15]. 1960년 지능형시스템 기술에 대한 최초특허가 등록된 이후 2008년까지 출원, 등록된 총 특허수는 모두 1537건이었다. 다음 표는 연도별 특허 건수를 나타내고 있다.

표 1. 연도별 특허건수
Table 1. # of patent by year

연도	특허건수	연도	특허건수	연도	특허건수
1960	1	1985	6	1997	129
1974	2	1986	7	1998	128
1975	2	1987	5	1999	148
1976	1	1988	14	2000	140
1977	2	1989	14	2001	166
1978	4	1990	17	2002	109
1979	3	1991	18	2003	108
1980	3	1992	18	2004	101
1981	4	1993	35	2005	72
1982	1	1994	41	2006	55
1983	2	1995	57	2007	24
1984	4	1996	92	2008	4

다음 그림은 본격적으로 지능형시스템에 대한 기술특허가 나타난 1974년부터 2006년까지의 연도별 특허출원건수의 추이를 나타내고 있다.

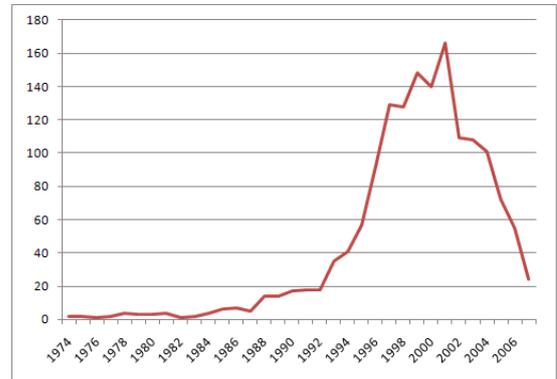


그림 2. 연도별 특허건수
Fig. 2. # of patent by year

하나의 특허가 출원되어 심사과정을 거쳐 최종적으로 등록이 이루어지기까지 길게는 수년의 기간이 소요된다. 본 논문에서는 2010년 5월 17일 현재까지 최종 등록된 특허만을 대상으로 하기 때문에 2007년 전, 후부터 출원된 특허들 중에서 아직 등록까지 이루어지지 않은 것들이 상당수 있을 것으로 판단되어 2007년 이후의 특허는 연도별 추이에서는 제외하였다. 위 그림을 통하여 1990년대 중반부터 증가하던 특허출원은 2000년대 초반에 가장 활발한 등록을 보인 이후 점차 감소하고 있음을 알 수 있다. 1980년대 후반부터 1990년대 중반까지 전 세계적으로 큰 관심을 받았고 이후 다소 침체되었던 인공지능에 대한 연구, 개발과 연관성이 있을 것으로 판단된다. 본 연구에서는 2000년까지의 특허데이터(training data)를 이용하여 기술예측모형을 만들고, 2000년 이후의 특허데이터(test data)를 이용하여 예측모형의 성능을 확인하였다. 다음 표는 본 논문의 실험에 사용된 특허데이터이다.

표 2. 학습데이터와 검증데이터
Table 2. Training and test data

Data	출원연도	등록 특허건수
Training	1960 - 2000	898
Test	2001 - 2008	639

학습데이터에 속한 898건의 특허문서를 이용하여 지능형시스템에 관한 공백기술의 예측모형을 구축하였다. 우선 각 특허문서의 초록을 이용하여 특허-키워드행렬을 구축하였다. R-Project의 'tm' 패키지에서 제공하는 텍스트마이닝 기법[18]을 이용하여 (898×6421)의 차원을 갖는 특허-키워드행렬을 구하였다. 'tm' 패키지는 공개된 통계소프트웨어 언어인 R 기반의 텍스트마이닝 모듈(module)이다. 윈도우와 유닉스 환경 모두에서 실행되고 계산속도도 다른 통계 패키지들에 비해 빠른 장점을 가지고 있다. 관측치(observation)에 해당되는 특허문서가 898개이고 변수(variable)인 키워드가 6421개인 행렬구조이다. 이 행렬데이터로부터 분석을 위한 특징추출(feature selection)을 수행한다. 앞서서도 언급한 것처럼 주성분에 의한 차원축소는 관측치의 수(n)가 변수의 수(p)보다 훨씬 커야만($n \gg p$)

가능하다. 그러므로 기존의 특징추출 방법을 보통의 특허데이터에 곧바로 적용하기에는 어려움이 있다. 이와 같은 문제점을 해결하기 위하여 본 논문에서는 SVD에 의한 주성분분석을 수행하여 모두 898개의 특징을 구하였다. 특히 본 논문에서는 다음 그림과 같이 가장 큰 설명력을 갖는 상위 10개의 특징을 사용하였다.

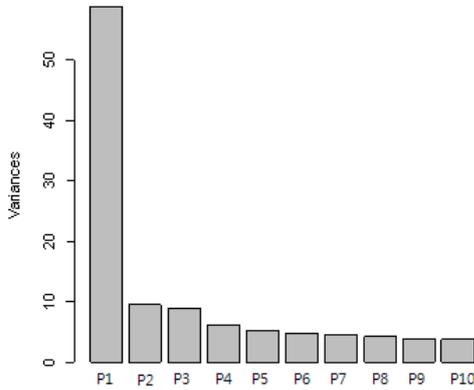


그림 3. 상위 10개 특징의 분산
Fig. 3. Variances of Top 10 features

위 그림은 상위 10개 특징들의 변동(설명력)을 나타내고 있다. P1-P10은 모두 10개의 SVD 기반의 주성분들이다. P1이 전체 데이터의 약 60%를 설명하고 있는 제1주성분이고 P10은 상위 10번째 주성분으로 약 5%의 설명력을 가지고 있다. 즉, 그림 3에서 가로축은 각각의 특징을 나타내고 세로축은 각 특징에 해당하는 설명력을 나타낸다. 예를 들어 두 번째로 큰 설명력을 갖는 두 번째 특징(P2)의 분산은 약 10으로 전체 데이터의 약 10%를 설명하고 있다. 상위 10개의 특징이 전체의 95% 이상을 설명하고 있다. 다음으로 10개의 추출된 특징을 군집변수로 이용하여 898개의 특허에 대한 군집화를 수행한다. 군집화를 수행하기에 앞서 먼저 최적의 군집수를 결정해야 한다. 물론 분석가가 경험에 의해 주관적으로 군집수를 결정할 수도 있지만 군집화의 성능향상을 위하여 보다 객관적인 군집수 결정이 요구된다. 본 논문에서는 대표적인 객관적인 군집화측도들 중의 하나인 BIC를 이용하였다. 가장 작은 BIC값에서 최적의 군집수가 결정된다. 다음 그림은 군집수에 따른 BIC 값을 나타내고 있다.

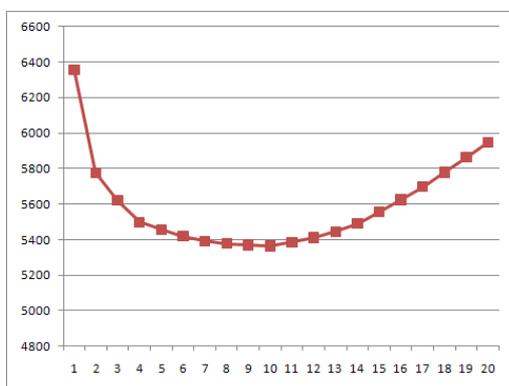


그림 4. 군집수에 따른 BIC값
Fig. 4. BIC values according to # of clusters

가장 작은 BIC값을 갖는 결과의 군집수는 10이었다. 이 결과를 통하여 지능형시스템에 관한 세부기술은 대략 10가지 정도로 나눌 수 있음을 알 수 있다. 물론 이 분야의 전문가 집단으로부터 기술분류를 통하여 세부기술을 나눌 수도 있지만 본 논문에서는 이 분야의 특허출원 및 등록의 결과를 통하여 객관적으로 이 작업을 수행하였다. 이 결과를 이용하여 군집수를 알 경우에 우수한 군집화 결과를 제공하는 K-means 군집화 기법을 사용한다. 본 논문에서는 K가 10인 K-means 군집화[21],[22]를 수용하여 학습데이터에 속한 898개의 특허문서들에 대한 군집화를 수행하였다. 다음은 K-means 군집화 결과로부터 10개의 각 군집에 속한 특허문서들의 할당결과를 나타내고 있다.

표 3. K-means 군집화 결과 (K=10)
Table 3. Result of K-means clustering (K=10)

군집	소속특허 일련번호	특허 수 (%)
1	123, 177, ..., 997	14 (1.56)
2	394	1 (0.11)
3	2, 7, ..., 1370	297 (33.07)
4	8, 60, ..., 1360	12 (1.34)
5	4, 13, ..., 1367	48 (5.35)
6	98, 104, ..., 1165	17 (1.89)
7	6, 12, ..., 1365	246 (27.39)
8	28, 57, ..., 1364	130 (14.48)
9	56, 108, 1326	41 (4.57)
10	3, 20, ..., 1363	92 (10.24)
합계		898 (100)

군집3과 7은 각각 전체 기술특허의 33.07%와 27.39%를 차지하고 있다. 이들 2개의 세부기술이 지능형시스템의 10가지 세부기술들의 60% 이상을 출원, 등록하고 있다. 때문에 이 기술분야는 활발한 연구개발이 이루어지고 있을 뿐만 아니라 중복연구와 투자가 이루어 질수 있는 위험이 존재하는 세부기술 분야들이다. 군집5, 9, 그리고 10에 해당되는 세부기술들은 비교적 활발한 연구개발이 이루어져오고 있음을 알 수 있다. 군집2는 1개의 이질적인 기술로 이루어진 특허군집이다. 1999년에 등록된 이 특허의 명칭은 'System method for facilitating intelligent network call processing using a call context database'이다. 1999년 이전에도 그리고 2000년 이후에도 이와 같은 부류의 기술특허가 출원, 등록되지 않고 있기 때문에 본 논문에서 이 기술은 지능형시스템의 공백기술예측을 위하여 고려될 필요가 없다고 판단하였다. 군집1, 4, 그리고 6은 각각 전체 출원, 등록특허의 1.6%, 1.3%, 그리고 1.9%에 해당된다. 이들 3가지 세부기술들로부터 지능형시스템에 관한 공백기술을 찾기 위하여 본 논문에서는 다음 그림과 같이 3개의 각 군집에 해당하는 특허의 연도별 등록건수를 표와 그림을 통하여 정리하였다.

지능형시스템 분야에서 군집6에 해당하는 기술특허는 1978년부터 매년 꾸준히 관련 기술특허가 출원, 등록되고 있다. 이에 비해 군집1은 1979년에 1건이 등록되고 이후 특허가 등록되지 않다가 1995년 이후 상대적으로 많은 특허가 등록되고 있다. 군집4는 군집1과 6과는 달리 비교적 최근인 1990년에 처음으로 이 군집에 해당되는 세부기술이

특히 등록되고 최근까지 기술특허가 지속적으로 이루어지고 있다. 이와 같은 연도별 특허 등록건수를 고려하여 본 논문에서는 군집1과 4가 지능형시스템에서 필요한 세부 공백기술로 판정하였다. 일단 공백기술에 해당되는 기술특허의 군집이 결정되었기 때문에 다음으로 각 군집에 소속된 특허들로부터 키워드를 추출하고 이를 바탕으로 각 군집에 해당되는 세부기술을 정의하였다. 다음 표는 군집1과 4를 포함한 전체 군집들에 속한 특허문서들로부터 추출한 상위 10개의 키워드이다. 다만 군집2는 1개의 특허로 이루어진 기술군집이기 때문에 제외하였다

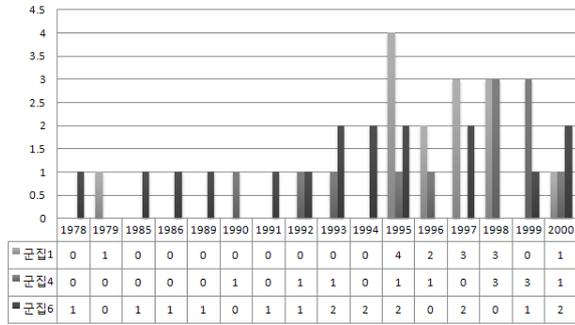


그림 5. 3개 그룹에 대한 연도별 특허 등록건수
Fig. 5. # of registered patents by year (group 1, 2, 3)

표 4. 군집1과 4에 속한 특허문서들로부터의 상위 10개 키워드
Table 4. Top 10 keywords from patent documents of group 1 and 4

군집	상위 10개 키워드
1	bus, component, local, pci, commands, scsi, hub, power, serial, connected
3	terminal, synchronous, radio, circuitry, message, peripheral, fuel, fuzzy, gate, lean
4	network, fuzzy, traffic, used, message, beam, optical, circuitry, general, energy
5	wave, cache, microingredient, service, backup, configuration, ifd, frequency, handset, call
6	host, pool, device, cpu, instruction, memory, holding, document, printer, speed
7	synchronous, address, switch, battery, cooking, intelligent, serial, electronic, brake, bridge
8	designated, recipient, mechanism, optical, section, liquid, transport, tcap, asset, tangible
9	brew, distributed, means, transfer, storage, transmission, property, cache, roadway, telescope
10	component, software, audio, query, message, measurement, buffer, flash, vehicle, reminder

군집1에 속하는 지능형시스템에 대한 세부기술은 이 군집에 속하는 10개의 키워드로부터 정의될 수 있다. 군집4도 같은 방법으로 정의될 수 있다. 본 논문에서는 군집1이 나타내는 공백기술은 지능형 입출력(I/O)에 관한 세부기술로 판단하였다. 그리고 군집4는 지능형 통신에 관한 세부기술

을 공백기술로 정의하였다. 이와 같은 결과를 이용하여 2001년부터 2008년에 출원, 등록된 639건의 지능형시스템에 대한 특허 검증데이터로부터 군집1과 4에 해당되는 기술 특허를 모두 검색하였다. 다음 표는 이에 대한 결과이다.

표 5. 검증데이터에서 군집1과 4에 속한 특허건수
Table 5. # of patents of group 1 and 4 in test data

군집	특허수	특허비율(%) (전체 639건)
1	7	1.10
4	72	11.27

이전의 표3에서 1960년부터 2000년까지의 총 898건의 출원, 등록 특허 중에서 군집1과 4에 속한 특허의 비율은 각각 1.56%와 1.34%였다. 군집1과 4는 본 논문에서 지능형시스템에 관한 세부공백기술로 판정된 기술군집이다. 위 표의 결과를 통하여 2개의 공백기술 군집 중에서 군집4는 2001년 이후 8배 이상의 특허건수 증가가 이루어진 것을 알 수 있었다. 하지만 군집1은 1.10%로 이전의 1.56%와 별 차이를 보이지 않고 있다. 따라서 군집1은 공백기술이기 보다는 지능형시스템 분야에서 지속적으로 1~2%의 점유율을 차지하는 세부기술로 고려될 수 있다. 하지만 군집4는 분석시점인 2000년에 향후 많은 기술개발이 필요하게 될 공백기술이었음이 확인되었다.

5. 결론 및 향후연구

본 논문에서는 특허문서를 이용한 공백기술예측에 대한 방법을 제안하고, 이를 지능형시스템 분야에 적용하였다. 학교, 기업, 그리고 정부의 연구기관들이 서로 중복된 연구를 수행할 수도 있고, 불필요한 기술에 대하여 시간과 비용을 낭비할 수도 있다. 본 연구에서 제안된 방법을 이용하여 특허데이터를 분석하고, 공백기술을 예측하여 이를 바탕으로 향후 지능형시스템 분야에서 필요하게 될 세부기술에 대한 계획적인 연구, 개발이 이루어 질 수 있을 것이다. 물론 본 연구에서 제시한 방법은 지능형시스템분야 이외의 다양한 기술분야에도 적용될 수 있다. 하지만 본 연구의 실험결과에서 지능형시스템 분야의 세부 공백기술로 2개의 기술군집을 예측했지만 이들 중 하나는 공백기술이 아닌 것으로 실험결과 확인되었다. 이와 같은 문제를 해결하기 위해서 좀 더 정교한 분석기법에 대한 연구와 함께 분석결과에 대한 일반화된 해석을 위하여 이 분야의 전문가집단에 대한 폭넓은 참여도 필요할 것이다.

본 논문에서는 지능형시스템에 대한 연구결과로서 지금까지 출원, 등록된 미국특허만을 이용하였다. USPTO의 특허데이터베이스에는 미국의 학교, 기업 및 정부의 모든 연구기관의 연구결과와 미국시장을 겨냥한 전 세계의 연구기관으로부터 출원된 결과를 포함하고 있지만 지능형시스템에 대한 모든 연구결과를 포함한 것은 아니다. 앞으로 미국 특허청 뿐만 아니라 유럽특허청, 일본특허청, 그리고 대한민국특허청으로부터의 특허데이터와 IEEE를 포함한 주요 논문지에 발표된 논문까지 포함한 광범위한 연구개발 결과를 이용한 기술예측이 이루어진다면 지능형시스템 기술에 대한 좀 더 정확한 기술예측 결과가 제공될 수 있을 것이다. 이 부분과 더 발전된 새로운 기술예측 방법에 대한 연구는

향후 과제로 남긴다.

참 고 문 헌

- [1] Han, J., Kamber, M., *Data Mining Concepts and Techniques*, Morgan Kaufmann, 2001.
- [2] 나카야마 노부히로, 특허법, 법문사, 2001.
- [3] 황종환, 특허법, 한빛지적소유권센터, 2001.
- [4] 특허청 정보기획팀, 한국발명진흥회 정보활용지원팀, 특허와 정보분석(개정판), 성민, 2007.
- [5] 남영준, 정의섭 "인용정보를 이용한 신 특허지수 개발에 관한 연구," *정보관리학회지*, vol. 23, no. 1, pp. 221-241, 2006.
- [6] 유선희, "특허인용 분석을 통한 기술수명예측모델 개발에 관한 연구," *정보관리연구*, vol. 35, no. 1, pp. 93-112, 2004.
- [7] 전성해, 박상성, 신영근, 장동식, 정호석, "자기조직화지도와 매트릭스분석을 이용한 특허분석시스템의 공백기술 예측," *한국콘텐츠학회논문지*, vol. 10, no. 2, pp. 462-480, 2010.
- [8] 전성해, 엄대호, "특허와 통계학, 그 연결은?" *한국통계학회논문집*, vol. 17, no. 2, pp. 205-222, 2010.
- [9] 제대식, 이은철, 윤국섭 역, *지식경영과 특허전략*, 세종서적, 2000.
- [10] Lee, S., Yoon, B., Park, Y., "An approach to discovering new technology opportunities: Keyword-based patent map approach," *Technovation*, vol. 29, pp. 481-497, 2009.
- [11] Wang, P., Cockburn, I. M., Puterman, M. L., "Analysis of Patent Data-A Mixed Poisson Regression Model Approach," *Journal of Business & Economic Statistics*, vol. 16, no. 1, pp. 27-41, 1998.
- [12] Yoon, B., Park, Y., "Development of New Technology Forecasting Algorithm: Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information," *IEEE Transactions on Engineering Management*, vol. 54, no. 3, pp. 588-599, 2007.
- [13] Yoon, B., Lee, S., "Patent analysis for technology forecasting: Sector-specific applications," *Proceeding of IEEE International Conference on Engineering Management*, pp. 1-5, 2008.
- [14] 특허정보검색서비스, <http://www.kipris.or.kr>
- [15] United States Patent and Trademark Office, www.uspto.gov
- [16] Fattori, M., Pedrazzi, G., Turra, R., "Text mining applied to patent mapping: a practical business case," *World Patent Information*, vol. 25, pp. 335 - 342, 2003.
- [17] Feinerer, I., Hornik, K., Meyer, D., "Text Mining Infrastructure in R," *Journal of Statistical Software*, vol. 25, iss. 5, pp. 1-54, 2008.
- [18] R-Project www.r-project.org
- [19] Hair, J. F., Black, B., Babin, B., Anderson, R. E., *Multivariate Data Analysis*, Prentice Hall, 1992.
- [20] Golub, G. H., Reinsch, C., "Singular value decomposition and least squares solutions," *Numerische Mathematik* vol. 14, no. 5, pp. 403-420, 1970.
- [21] Chen, X., Yin, W., Tu, P., Zhang, H., "Weighted k-Means Algorithm Based Text Clustering," *Proceedings of International Symposium on Information Engineering and Electronic Commerce*, pp. 51-55, 2009.
- [22] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., Wu, A. Y., "An efficient k-means clustering algorithm: Analysis and implementation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 881 - 892, 2002.

저 자 소 개



전성해(Sunghae Jun)

1993년 : 인하대 통계학과(학사)
 1996년 : 인하대 통계학과(이학석사)
 2001년 : 인하대 통계학과(이학박사)
 2007년 : 서강대학교 컴퓨터공학과
 (공학박사)
 2003년~현재 : 청주대학교 바이오정보통계학과 부교수

관심분야 : 기술경영, 인공지능, 데이터마이닝
 Phone : 043-229-8205
 Fax : 043-229-8432
 E-mail : shjun@cju.ac.kr