



## 베이지안 이산모형을 이용한 기술예측

### Technology Forecasting using Bayesian Discrete Model

전성해\*

Sunghae Jun

\*청주대학교 통계학과

\*Department of Statistics, Cheongju University

#### 요약

기술예측은 과거부터 현재까지의 기술개발 결과를 수집, 분석하여 특정 기술의 미래 추세 및 상태를 예측하는 것이다. 일반적으로 특허는 현재까지의 기술개발 결과를 가장 잘 가지고 있다. 왜냐하면 특허에 포함된 세부 기술은 일정기간 동안 배타적 권리가 법에 의해 보장되기 때문이다. 따라서 특허 데이터의 분석을 이용한 기술예측의 다양한 연구가 진행되었다. 특허문서의 분석을 위하여 널리 사용되는 특허 키워드 데이터는 주로 기술키워드에 대한 빈도 값으로 이루어진다. 기존의 많은 특허분석에서는 회귀분석, 박스-젠킨스 모형 등 연속형 데이터분석 기법이 적용하였다. 하지만 빈도 데이터는 이산형 데이터이기 때문에 이산형 데이터분석 방법을 사용해야 한다. 본 연구에서는 이와 같은 문제점을 해결하기 위하여 베이지안 포아송 이산모형을 이용한 특허분석 방법을 제안한다. 연구방법의 성능평가를 위하여 지금까지 출원, 등록된 애플의 전체특허를 분석하여 향후 기술을 예측하는 사례분석을 수행한다.

키워드 : 특허분석, 기술예측, 베이지안 포아송모형, 텍스트마이닝, 애플특허

#### Abstract

Technology forecasting is predict future trend and state of technology by analyzing the results so far of developing technology. In general, a patent has novel information about the result of developed technology, because the exclusive right of technology included in patent is protected for a time period by patent law. So many studies on the technology forecasting using patent data analysis has been performed. The patent keyword data widely used in patent analysis consist of occurred frequency of the keyword. In most previous researches, the continuous data analyses such as regression or Box-Jenkins Models were applied to the patent keyword data. But, we have to apply the analytical methods of discrete data for patent keyword analysis because the keyword data is discrete. To solve this problem, we propose a patent analysis methodology using Bayesian Poisson discrete model. To verify the performance of our research, we carry out a case study by analyzing the patent documents applied by Apple until now.

Key Words : Patent Analysis, Technology Forecasting, Bayesian Poisson Model, Text Mining, Apple Patents

Received: Feb. 1, 2017  
Revised: Feb. 26, 2017  
Accepted: Feb. 26, 2017  
† Corresponding authors  
shjun@cju.ac.kr

## 1. 서론

특정 기술 분야에 대한 기술예측(technology forecasting)은 미래에 출현 가능성이 큰 세부기술의 추세 및 상태를 예측하는 것이다 [1]. 기술예측과 관련된 그 동안의 연구결과에서 기술예측은 크게 3가지 접근방법에 의해 이루어졌다 [2]. 통계학과 기계학습(machine learning) 알고리즘을 이용한 정량적(quantitative) 기술예측, 델파이(Delphi)와 시나리오분석(scenario analysis) 등에 의한 정성적(qualitative) 기술예측, 그리고 정량적 방법과 정성적 방법을 결합한 통합(integrated) 기술예측이 사용되고 있다 [3-5]. 전문가의 기술지식과 주관적 판단에 의존하는 정성적 기술예측은 구성되는 전문가 집단의 능력에 따라 예측결과가 다르게 나타날 수 있기 때문에 최근의 기술예측은 통계적 기법을 이용하여 특허 빅데이터를 분석하는 정량적 접근방법에 의존하고 있다 [6-10]. 즉 전 세계 특허청 및 관련 기관의 특허데이터베이스로부터 목표기술에 대한 특허문서를 검색하고 텍스트 마이닝(text mining)의 전처리(preprocessing) 과정을 거쳐서 정형화된 데이터(structured data)를 구축하고, 이를 이용하여 통계학 또는 기계학습 기법에 의해 특허분석이 이루어진다. 사회네트워크분석(social network analysis), 신경망, 회귀분석, 군집분석, 시계열분석 등 다양한 분석기법들이 사용되었다. 정형화된 특허 데이터는 개별 특허문서로 이루어진 행(row)과 기술 키워드인 열(column)로

이 논문은 2016학년도에 청주대학교 산업과 학연구소가 지원한 학술연구조성비(특별연구과제)에 의해 연구되었음.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

구성된 행렬구조로 구축된다 [8, 11-12]. 이 행렬의 각 원소는 특정 키워드가 각 특허문서에 나타난 빈도수(frequency)이다. 따라서 각 기술키워드 간의 연관성 분석을 통하여 특정 기술 분야에서 개발된 세부 기술 간의 관계를 파악한다. 현재까지 대부분의 특허분석에 사용된 데이터분석 기법은 정형화된 특허데이터 행렬의 빈도수를 연속형(continuous) 데이터 형태로 가정하고 기술예측을 위한 분석이 이루어졌다. 하지만 빈도 데이터는 이산형(discrete) 데이터이기 때문에 이산형 특허분석이 필요하게 되었다. 또한 특허데이터 행렬의 기술키워드는 단순한 변수로 취급하기에는 각 키워드가 포함하고 있는 기술적 의미가 크기 때문에 이에 대한 고려가 구축되는 분석모형에서 이루어져야 한다. 즉 각 키워드가 포함하는 사전정보(prior)를 모형에 반영하는 작업이 필요하게 된다. 이와 같은 문제점을 해결하기 위하여 본 논문에서는 베이지안(Bayesian) 기반의 이산모형을 이용한 특허분석을 제안한다. 이를 통하여 향후 기술에 대한 효율적인 예측이 가능하도록 한다. 특히 사례연구를 위하여 대표적인 기술혁신 기업인 애플(Apple)의 전체특허를 분석하여 애플의 향후 기술에 대한 예측을 수행한다. 다음 장에서는 베이지안 이산모형을 이용한 기술예측 방법을 제시한다. 애플의 특허데이터를 이용한 사례연구는 3장에서 다루고, 마지막 장에서는 본 논문의 결론과 향후 연구과제에 대하여 논한다.

## 2 기술예측을 위한 베이지안 포아송 모형

자연과학이나 공학뿐만 아니라 사회과학 분야에서도 베이지안 기법들이 널리 사용되고 있다 [13]. 베이지안 통계학은 다음과 같이 사전확률함수(prior function), 가능도함수(likelihood function), 그리고 사후확률함수(posterior function)가 지속적인 학습을 통하여 갱신되는 구조를 갖는다 [14].

$$Posterior \propto Likelihood \times Prior \quad (1)$$

즉 사후함수의 확률은 관측된 데이터에 의한 가능도함수 값에 사전확률함수 값을 곱하여 계산된다. 새로운 데이터가 관측되면 현재 계산된 확률값은 다시 사전확률값으로 사용된다. 따라서 새로운 데이터가 수집될 때마다 사후확률은 갱신(update)된다. 이와 같은 학습(learning) 과정을 통하여 베이지안은 비베이지안(non-Bayesian) 분석기법에 비해 효과적인 분석능력을 갖는다 [14]. 그러나 비베이지안에 비해 베이지안에서는 더 많은 계산량이 요구되면서 이를 해결할 수 있는 다양한 방법이 제안되었다 [15-17]. 특히 다양한 예측모형을 구축하기 위하여 대부분의 베이지안 접근방법은 마코프체인 몬테칼로(Markov Chain Monte Carlo, MCMC) 기법을 사용한다 [13]. 통계적 추론(statistical inference)을 위하여 베이지안

모형은 이론적이면서 동시에 실제적인 기반을 모두 갖추고 있다 [16,18,19]. 또한 사후 모의실험(posterior simulation)을 이용한 베이지안 추론을 통하여 베이지안 이산모형을 구축할 수 있다 [13]. 본 논문은 베이지안 포아송 변화점(Bayesian Poisson Change Point, BPCP) 분석을 이용한 기술예측 방법을 제안한다. 시계열(time series) 데이터의 적합을 위하여 사용되는 기존의 변화점(change point) 통계모형과 마찬가지로 BPCP도 시간에 따른 변수의 추세를 효율적으로 모형화 한다 [20,21]. 키워드에 대한 베이지안 포아송 모형은 기본적으로 다음과 같은 베이지안 회귀모형을 기반으로 한다.

$$E(y_i|t_i) = \exp(\beta_0 + \beta_1 t_i + \epsilon_i), \quad i = 1, 2, \dots, n \quad (2)$$

오차  $\epsilon_i$ 는 평균이 0이고 분산이  $\sigma^2$ 인 정규분포를 따른다. 본 논문에서  $y_i$ 는 키워드의 빈도수를 나타내고  $t_i$ 는 연도를 나타낸다.  $\beta_0$ 는 모형의 절편을 나타내고  $\beta_1$ 은  $y$ 에 대한  $t$ 의 회귀계수(regression parameter)를 나타낸다. 즉  $\beta_1$ 이 통계적으로 유의하게 되면 해당 키워드는 연도에 의해 통계적으로 유의함을 확인할 수 있다. 식 (2)의 모형은 베이지안 추론에 따른 가능도함수와 다음의 사전확률함수를 이용하여 사후확률모형을 구축한다.

$$P(\beta_0, \beta_1, \sigma) = p_1(\beta_0)p_2(\beta_1)p_3(\sigma) \quad (3)$$

본 연구에서는 일반적인 베이지안 추론의 기본설정에 따라  $p_1$ 과  $p_2$ 는 일정한 상수(constant)에 비례하게 설정하고,  $p_3$ 는  $\sigma$ 에 반비례하게 설정한다 [14]. 베이지안 포아송 모형에서  $y$ 는 다음의 포아송 분포를 따른다.

$$P(Y = y|m) = \frac{e^{-m} m^y}{y!}, \quad y = 0, 1, 2, \dots \quad (4)$$

포아송 확률변수는 단위 시간동안 특정 사건의 발생 빈도수를 나타내기 때문에 제안방법에서 시간에 따른 특허 키워드의 빈도수에 대한 모형으로 적절하게 사용될 수 있다. 식 (4)에서 포아송 분포의 모수  $m$ 은 평균발생회수를 나타낸다. 베이지안 포아송 모형을 위한 모수  $m$ 은 다음과 같은 감마(gamma) 분포를 따른다.

$$f(m|a, b) = \frac{1}{\Gamma(a)b^a} m^{a-1} e^{-m/b}, \quad 0 < m < \infty, a > 0, b > 0 \quad (5)$$

식(5)에서  $\Gamma(\cdot)$ 는  $\Gamma(n) = (n-1)!$ ( $n > 0$ )인 감마함수(gamma function)이고,  $a$ 와  $b$ 는 각각 형상(shape)과 척도(scale)를 나타내는 모수이다. 이와 같은 과정에 의해 본 논문의 베이지안 포아송 이산모형이 구축된다. [13, 20]. 이 모형을 통하여 특허문서로부터

추출된 키워드에 대한 분석을 통해 각 기술키워드의 연도별 추세를 예측한다. 다음 그림은 연도별 기술 키워드 분석을 위한 베이지안 포아송 모형을 계층적으로 설명하고 있다.

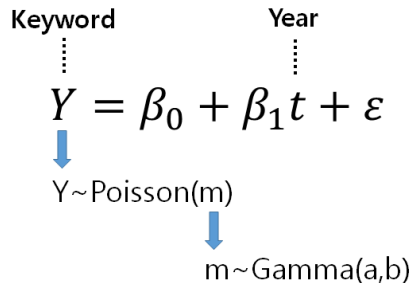


그림 1. 연도별 키워드 분석을 위한 베이지안 포아송 모형  
Fig. 1. Bayesian Poisson model for keyword analysis by year

기존의 선형 모형에 적용된 확률분포의 모수에 대한 사전확률분포를 추가로 사용함으로써 모형의 성능을 향상시킨다. 따라서 본 연구에서는 목표 기술분야의 선정부터 특허문서의 수집 및 전처리, 분석을 통해 최종적으로 기술의 연구개발 기획에 이르는 전체 과정을 다음과 같이 7단계로 나눈다.

- 단계 1: 여러 특허 데이터베이스로부터 주어진 기술 분야에 대한 특허 문서를 검색한다.
- 단계 2: 수집된 특허 문서들 중에서 최종적으로 분석에 합당한 유효특허를 필터링한다.
- 단계 3: 텍스트 마이닝을 이용한 전처리 과정을 통하여 통계분석이 가능한 구조화된 특허 데이터를 구축한다.
- 단계 4: 기술 키워드에 대한 상자그림을 이용하여 시간에 따른 각 기술의 추세를 파악한다.

단계 5: 베이지안 포아송 이산모형을 이용하여 시간적 영향력이 큰 기술키워드를 찾아낸다.

단계 6: 상자그림과 베이지안 분석의 결과를 이용하여 향후 기술경향을 예측한다.

단계 7: 기술예측 결과를 이용하여 기업의 연구개발 계획을 세운다.

전체적으로 기술예측을 위한 통계모형의 구축에 사용되는 특허 데이터는 텍스트 마이닝의 전처리 과정을 통하여 연도-키워드 행렬구조로 이루어진다. 즉 각 행은 연도를, 그리고 각 열은 기술 키워드를 나타낸다. 행렬의 원소는 연도에 따른 특정 키워드의 출현 빈도수를 나타낸다. 각 키워드의 빈도수는 베이지안 포아송 모형의 반응변수  $Y$ 가 되고 연도는 시간을 나타내는 설명변수  $t$ 가 된다. 제안 방법의 실제 적용과정을 보이기 위하여 다음 절에서 지금까지 애플이 출원 및 등록된 전체특허를 이용한 실험을 수행하였다.

### 3. 실험 및 결과

제안하는 기술예측 방법론의 실제 적용을 위하여 애플이 지금까지 출원, 등록된 전체특허를 수집하고 분석하였다. 삼성과 함께 전 세계 스마트폰 시장을 주도하는 애플은 지금까지 혁신적인 기술을 바탕으로 한 다양한 제품을 생산하고 있다. 따라서 본 논문에서는 전 세계 특허 데이터베이스로부터 애플이 출원, 등록된 전체 특허를 수집하였고 총 유효 특허수는 8,119건이었다 [22,23]. 특허 데이터의 분석을 위하여 본 실험에서는 R 데이터언어와 R이 제공하는 패키지를 사용하였다 [13, 24-26]. 텍스트 마이닝을 이용한 전처리 과정을 거쳐서 그림 2와 같은 연도별 키워드를 추출하였다.

	Year	device	data	system	user	media	display	computer	electronic	interface	
1	2012	7	0	0	0	7	0	0	0	0	
2	2010	0	1	5	0	0	0	0	0	0	
3	2010	1	0	0	1	0	0	0	1	0	
4	2012	3	0	0	0	0	5	0	1	0	
5	2010	3	0	2	0	0	0	0	4	0	
6	2012	0	0	5	0	0	0	0	0	0	
	information	image	memory	power	control	signal	video	plurality	audio	portable	application
1	0	0	0	0	0	0	0	0	0	6	0
2	0	0	3	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0
4	0	1	0	0	0	0	0	2	0	2	0
5	0	0	0	3	2	0	0	0	0	2	0
6	0	0	1	1	0	0	0	0	0	0	0
	invention	circuit	present	set	portion	digital	content	object	color	network	
1	0	0	0	0	0	0	4	0	0	4	
2	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	6	
4	0	0	0	0	1	0	0	0	0	0	
5	0	0	0	0	0	6	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	

그림 2. 애플특허의 연도-키워드 행렬  
Fig. 2. Year-keyword matrix of Apple patents

공통단어(common terms)인 ‘and’, ‘a’, ‘is’ 등의 단어와 애플의 기술과 관련이 없는 단어를 제외한 상위 단어들을 키워드로 선정하였다. 본 논문에서는 텍스트 마이닝을 위하여 R 데이터언어에 기반한 ‘tm’ 패키지를 이용하였다[24,25]. 연도-키워드 행렬 데이터의 첫 번째 열은 연도를 나타내고 2번째 열부터 마지막 열은 30개의 키워드를 나타낸다. 연도는 1977년부터 2012년까지의 값을 가지고 있다. 그림 3은 ‘Device’, ‘Data’, ‘System’, ‘User’, ‘Media’, 그리고 ‘Display’ 키워드들의 연도별 빈도수에 대한 상자그림(box plot)을 나타내고 있다. 본 연구에서는 특히 최고 빈도수의 추세와 상자의 크기를 통하여 연도별 키워드의 추세를 확인하였다.

‘Device’ 키워드는 1990년대 중 후반까지는 큰 빈도수를 나타내다가 이후 감소하였지만 최근에는 다시 증가하는 추세를

나타내고 있다. 이는 ‘Device’ 관련 기술의 개발 추세를 설명하고 있다. ‘Data’, ‘System’ 그리고 ‘Display’ 관련 기술들도 ‘Device’와 유사한 경향을 나타낸다. 반면에 ‘User’와 ‘Media’는 1990년대 이후 꾸준히 증가하고 있음을 알 수 있다. 또한 ‘Device’, ‘System’, 그리고 ‘Display’의 상자 크기가 다른 키워드들에 비해 크기 때문에 애플의 기술개발에 이들 키워드에 해당하는 기술의 영향력이 더 큼을 알 수 있다. 또한 ‘User’의 상자 크기가 과거에 비해 최근에 지속적으로 커지고 있음이 확인되었다. 이를 통하여 애플은 ‘User’와 관련된 기술에 대한 중요성이 증가하고 있음을 알 수 있다. 그림 4은 ‘Computer’, ‘Electronic’, ‘Interface’, ‘Information’, ‘Image’, 그리고 ‘Memory’의 연도별 키워드 빈도수를 나타내고 있다.

‘Computer’의 상자 크기가 다른 키워드들에 비해 상당히 큰

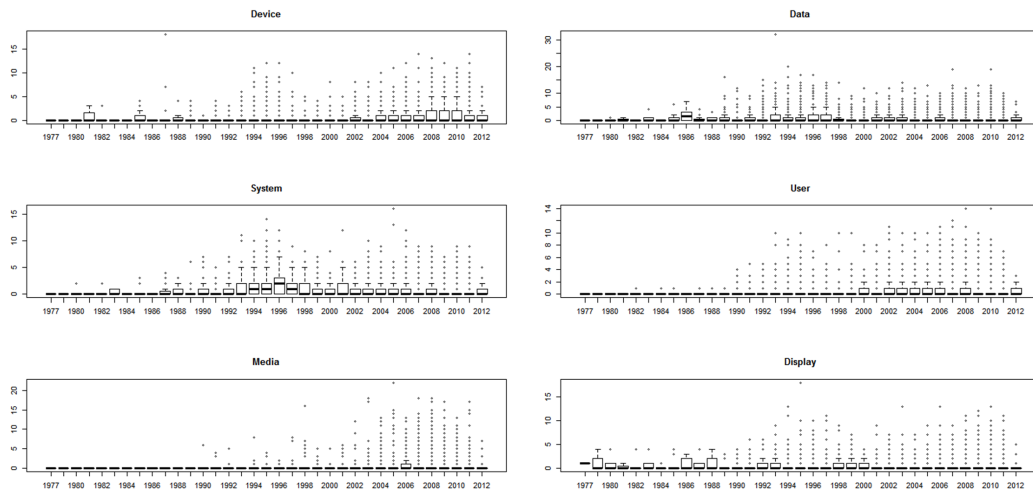


그림 3. 시간에 따른 키워드 상자그림 1: Device, Data, System, User, Media, Display  
Fig. 3. Keyword box-plot 1 by year: Device, Data, System, User, Media, Display

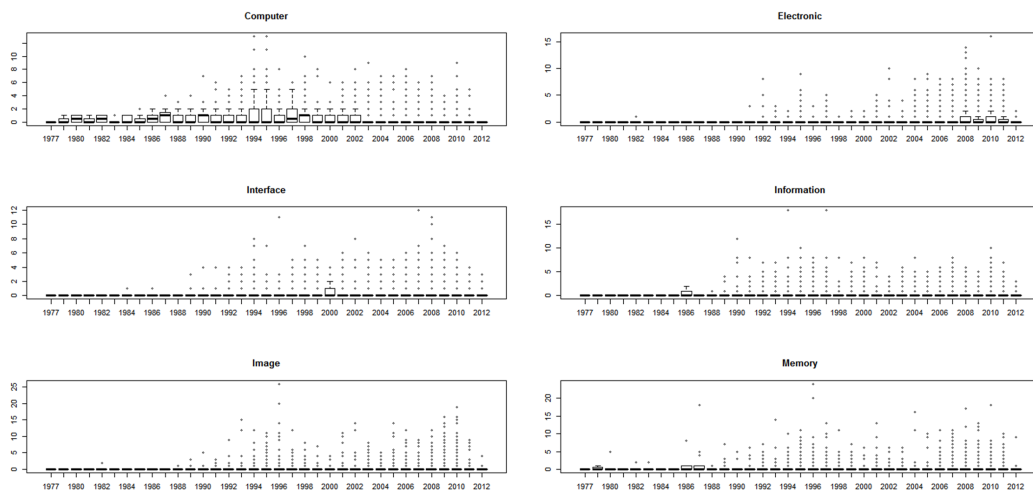


그림 4. 시간에 따른 키워드 상자그림 2: Computer, Electronic, Interface, Information, Image, Memory  
Fig. 4. Keyword box-plot 2 by year: Computer, Electronic, Interface, Information, Image, Memory

것을 확인할 수 있다. 최고 빈도수의 추세는 6개의 키워드들 간에 큰 차이는 나타내지 않고 있다. 그림 3의 'User'와 비슷하게 'Electronic' 키워드의 상자크기가 이전에 비해 최근에 커지는 경향을 나타내고 있다. 그림 5는 'Power', 'Control', 'Signal', 'Video', 'Plurality', 그리고 'Audio' 키워드들에 대한 연도별 추세를 나타내고 있다.

그림 5의 키워드들도 대부분 1990년대 중반까지 증가하다가 이후 감소를 보이다가 2000년대 들어서 다시 증가하는 경향을 나타내고 있다. 특히 'Video'에 해당하는 기술은 1980년대 중반 활발히 개발되었던 것을 알 수 있다. 그림 6는 'Portable', 'Application', 'Invention', 'Circuit', 'Present', 그리고 'Set'에 대한 연도별 키워드들의 시계열 추세를 나타내고 있다.

'Portable' 키워드는 이전에 비해 최근에 급속하게 증가하고

있음을 알 수 있다. 그림 7은 'Portion', 'Digital', 'Content', 'Object', 'Color', 그리고 'Network'에 대한 시계열 추세를 나타내고 있다.

'Content' 키워드는 2000년대 이후 급속하게 증가하고 있음을 알 수 있다. 나머지 5개의 키워드들의 추세는 애플의 여러 키워드들과 마찬가지로 1990년대 중반까지 증가하다가 이후 감소하고 2000년대 들어와서 증가하는 경향을 보이고 있다. 지금까지 상자그림을 통하여 개별 키워드에 대한 연도별 시계열 추세를 확인하고 애플의 기술개발에 영향력이 있는 키워드에 대하여 확인하였다. 이와 같은 시각화 결과를 바탕으로 본 논문의 베이지안 이산모형을 이용한 시계열 분석을 수행한다. 그림 8은 'Device' 키워드에 대한 분석결과를 나타낸다.

즉 반응변수는 'Device'가 되고 설명변수는 연도(Year)가 된다.

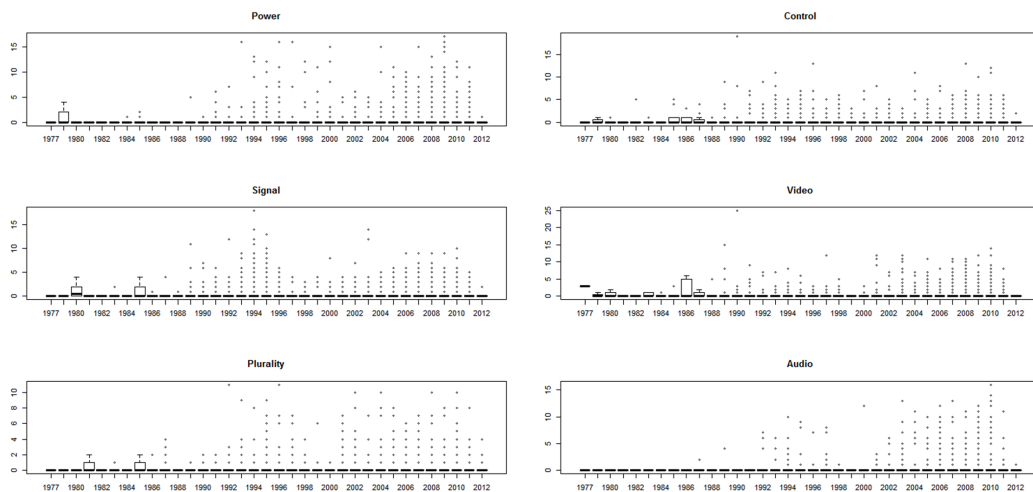


그림 5. 시간에 따른 키워드 상자그림 3: Power, Control, Signal, Video, Plurality, Audio  
Fig. 5. Keyword box-plot 3 by year: Power, Control, Signal, Video, Plurality, Audio

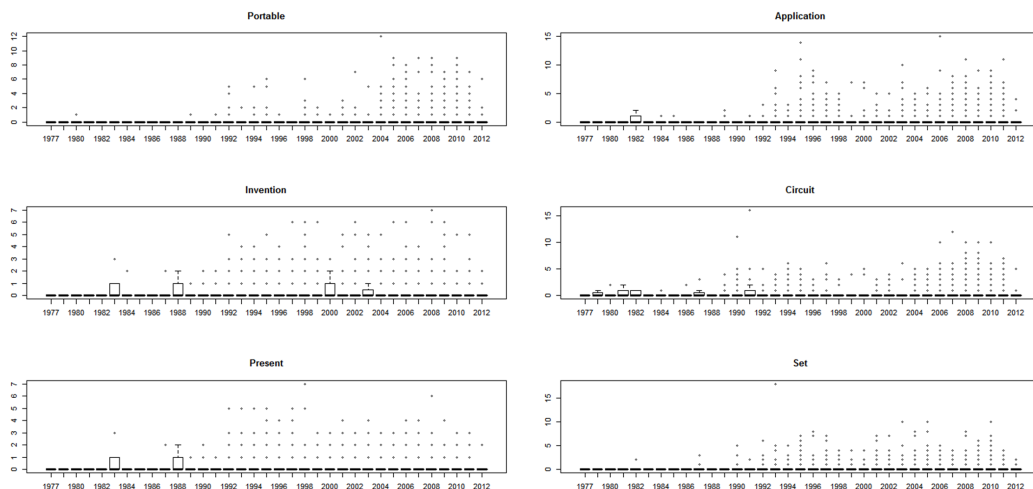


그림 6. 시간에 따른 키워드 상자그림 4: Portable, Application, Invention, Circuit, Present, Set  
Fig. 6. Keyword box-plot 4 by year: Portable, Application, Invention, Circuit, Present, Set



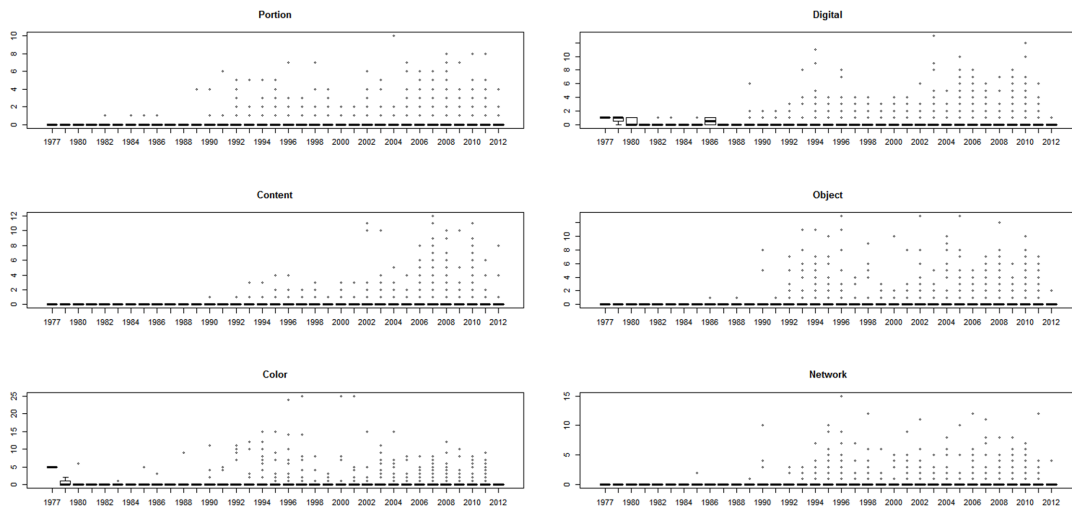


그림 7. 시간에 따른 키워드 상자그림 5: Portion, Digital, Content, Object, Color, Network  
 Fig. 7. Keyword box-plot 5 by year: Portion, Digital, Content, Object, Color, Network

그림 8의 상단의 2개 그림은 모형의 절편(intercept)에 대한 추이값(trace)과 모수에 대한 밀도 함수를 나타내고 있다. 상단의 첫 번째 그림에서 절편은 -1.0을 중심으로 움직이고 있지만 산포가 크지 않아 모형이 안정적임을 알 수 있다. 상단의 두 번째 그림은 절편에 대한 밀도함수를 나타내고 있다. 절편은 주어진 데이터의 척도(scale)에 따라 변화할 수 있다. 따라서 'Device'에 대한 추이와 모수가 중요하다. 하단의 첫 번째 그림을 통하여 'Device' 모수에 대한 추이를 확인할 수 있다. 하단의 두 번째 그림은 'Device'

키워드에 대한 모수값의 분포를 나타내고 있다. 0.4를 중심으로 좌, 우 대칭적인 정규분포를 나타내고 있음을 확인할 수 있다. 표 1은 모든 키워드에 대하여 베이지안 이산 시계열분석을 수행한 결과를 나타내고 있다.

각 키워드에 대한 추정된 모수(Beta), 표준편차(standard deviation, SD), 시계열 표준오차(time series standard error, TS-SE) 통하여 개별 키워드에 대한 영향력을 판단할 수 있다. 'Electronic' 키워드의 모수가 19.6293으로 가장 크게 나타났다. 모수값에 비례하여 'Electronic'의 표준편차와 시계열 표준오차도 크게 나타나고 있음을 알 수 있다. 즉 가장 큰 모수값과 산포를 갖는 'Electronic' 관련 기술이 애플의 전체 기술에 가장 큰 영향을 미침을 알 수 있다. 다음으로 'Computer', 'Color', 'Object', 그리고 'Network' 키워드들이 각각 1.2478, 0.9688, 0.9206, 0.9158의 모수값들로 애플의 기술에 영향을 미치고 있음을 알 수 있다. 즉 다음과 같은 순서에 의해 각 키워드에 해당하는 기술은 애플의 전체 기술에 영향을 미치게 된다; 1st-Electronic, 2nd-Computer, 3rd-Color, 4th-Object, 5th-Network, 6th-Present, 7th-Signal, 8th-Invention, 9th-Set, 10th-Digital, 11th-Circuit, 12th-Plurality, 13th-Control, 14th-Memory, 15th-Portion, 16th-Video, 17th-Application, 18th-Information, 19th-Image, 20th-Audio, 21st-System, 22nd-Data, 23rd-Power, 24th-Content, 25th-Interface, 26th-Display, 27th-Device, 28th-User, 29th-Media, 30th-Portable.

그림 3에서 그림 7까지의 결과에서 'User'와 'Electronic' 키워드의 상자의 크기가 최근에 와서 커지는 추세를 확인할 수 있었다. 이와 같은 결과는 앞으로 당분간 애플의 기술 및 제품은 'User'와 'Electronic' 관련 기술에 영향을 받는다고 할 수 있다. 또한 베이지안

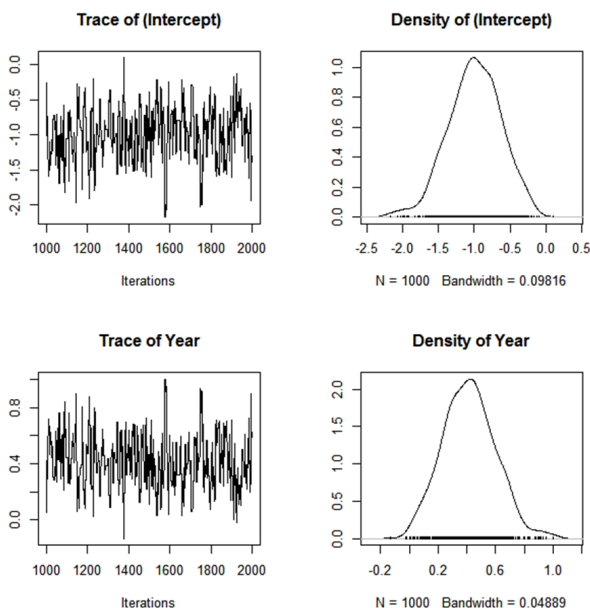


그림 8. 키워드 'Device'의 시계열 추세분석  
 Fig. 8. Time series trend analysis of keyword 'Device'

표 1. 추정된 모수와 오차  
Table 1. Estimated parameters and errors

Keyword	Beta	SD	TS-SE
Device	0,4161	0,1836	0,0166
Data	0,5415	0,1812	0,0164
System	0,5753	0,1840	0,0168
User	0,1825	0,1896	0,0171
Media	0,0737	0,2182	0,0217
Display	0,4711	0,1885	0,0172
Computer	1,2478	1,1204	0,3468
Electronic	19,6293	20,1674	18,0036
Interface	0,5142	0,1721	0,0144
Information	0,6725	0,1917	0,0177
Image	0,6019	0,1891	0,0174
Memory	0,7335	0,1878	0,0170
Power	0,5411	0,1804	0,0156
Control	0,7586	0,1894	0,0180
Signal	0,8552	0,1872	0,0175
Video	0,7308	0,1834	0,0167
Plurality	0,7739	0,1812	0,0161
Audio	0,5795	0,1872	0,0173
Portable	0,0362	1,8293	0,6286
Application	0,7236	0,1907	0,0173
Invention	0,8520	0,1908	0,0176
Circuit	0,7800	0,1888	0,0170
Present	0,8870	0,1835	0,0166
Set	0,8272	0,1943	0,0180
Portion	0,7327	0,1821	0,0170
Digital	0,8082	0,1941	0,0184
Content	0,5278	3,1953	1,4078
Object	0,9206	0,1878	0,0166
Color	0,9688	0,1924	0,0174
Network	0,9158	0,1856	0,0169

이산모형의 결과 ‘Electronic’, ‘Computer’, ‘Color’, ‘Object’, 그리고 ‘Network’에 해당하는 기술의 증가폭이 크다는 것을 알 수 있었다. 따라서 ‘User’, ‘Electronic’, ‘Electronic’, ‘Computer’, ‘Color’, 그리고 ‘Object’와 관련된 기술이 향후 애플의 기술 및 신상품 개발과정에서 중요하게 사용될 것으로 예측되었다.

#### 4. 결론 및 향후 연구

본 논문은 기술예측을 위한 특허 데이터의 분석방법에 대하여 연구하였다. 수집된 특허문서는 텍스트 마이닝의 전처리과정을 통하여 구조화된 데이터 형태로 변환되고 이를 통하여 통계분석이 가능하게 된다. 특허 키워드에 기반 한 대부분의 특허데이터는 특허(row)와 키워드(column), 그리고 빈도(element)로 이루어진 행렬형태를 띠고 있다. 기존에 특허분석은 회귀분석, 박스-젠킨스 시계열분석, 등 대부분 연속형 데이터분석 기법들에 의존하였다. 하지만 이산형인 빈도데이터에 대하여 연속형 분석기법을 적용하는 것보다 이산형 분석기법을 적용하는 것이 필요하였다. 이와 같은

문제를 해결하기 위하여 본 연구에서는 베이저안 이산모형을 적용한 특허분석을 수행하였다. 제안방법의 성능평가 및 실제 적용을 보이기 위하여 애플이 출원, 등록된 전체 특허문서를 수집하여 분석하였다. 제안된 베이저안 이산모형 이외에도 통계학에서는 다양한 이산 데이터의 분석기법들이 있다. 여러 이산모형의 결과들을 결합한 앙상블 전략을 통하여 타당성을 더 확보할 수 있는 모형개발에 대한 연구가 향후 과제에서 진행될 수 있다.

#### References

- [1] Roper, A. T., Cunningham, S. W., Porter, A. L., Mason, T. W., Rossini F. A, Banks J., *Forecasting and Management of Technology*, Hoboken, NJ, John Wiley & Sons, 2011.
- [2] Jun, S., Park, S., Jang, D., *Patent Analysis and Technology Forecasting*, Kyowoo, 2014.
- [3] Jun, S., Park, S., Jang, D., “Technology Forecasting using Matrix Map and Patent Clustering”, *Industrial Management & Data Systems*, Vol. 112, Iss. 5, pp. 786-807, 2012.
- [4] Keller, J., Gracht, H. A. V. D., “The influence of information and communication technology (ICT) on future foresight processes – Results from a Delphi survey”, *Technological Forecasting and Social Change*, Vol. 85, pp. 81-92, 2014.
- [5] Jun, S., Lee, S., Ryu, J., Park, S., “A novel method of IP R&D using patent analysis and expert survey,” *Queen Mary Journal of Intellectual Property*, Vol. 5, No. 4, pp. 474-494, 2015.
- [6] Jun, S., “A Big Data Learning for Patent Analysis”, *Journal of Korean Institute of Intelligent Systems*, Vol. 23, No. 5, pp. 406-411, 2013.
- [7] Jun, S., “A Big Data Preprocessing using Statistical Text Mining”, *Journal of Korean Institute of Intelligent Systems*, Vol. 25, No. 5, pp. 470-476, 2015.
- [8] Kim, J., Jun, S., “Graphical Causal Inference and Copula Regression Model for Apple Keywords by Text Mining”, *Advanced Engineering Informatics*, Vol. 29, Iss. 4, pp. 918-929, 2015.
- [9] Mishra, D., Gunasekaran, A., Childe, S. J., Papadopoulos, T., Dubey, R., Wamba, S., “Vision, applications and future challenges of Internet of Things: A bibliometric study of the recent literature”, *Industrial Management & Data Systems*, Vol. 116, No. 7, pp. 1331-1355, 2016.

- [10] Petruzzelli, A. M., Rotolo, D., Albino, V., “Determinants of patent citations in biotechnology: An analysis of patent influence across the industrial and organizational boundaries”, *Technological Forecasting and Social Change*, Vol. 91, pp. 208-221, 2015.
- [11] Kim, H., Kim, J., Lee, J., Park, S., Jang, D., “A Novel Methodology for Extracting Core Technology and Patents by IP Mining”, *Journal of Korean Institute of Intelligent Systems*, Vol. 25, No. 4, pp. 392-397, 2015.
- [12] Kim, J., Lee, J., Park, S., Jang, D., “Technology Strategy based on Patent analysis”, *Journal of Korean Institute of Intelligent Systems*, Vol. 26, No. 2, pp. 141-146, 2016.
- [13] Quinn, K. M., Martin, A. D., Park, J. H., MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, Vol. 42, No. 9, pp. 1-21, 2011.
- [14] Press, S. J., *Subjective and Objective Bayesian Statistics*, second edition, Hoboken, NJ, John Wiley & Sons, 2003.
- [15] Botcher S. G., Dethlefsen, C., “Learning Bayesian networks with R”, *International Workshop on Distributed Statistical Computing (DSC2003) Working Papers*, pp. 1-11, 2003.
- [16] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., *Bayesian Data Analysis*, Third Edition, Boca Raton, FL, Chapman & Hall/CRC Press, 2003.
- [17] Korb, K. B., Nicholson, A. E., *Bayesian artificial intelligence*, second edition, London, UK CRC press, 2011,
- [18] Jeffreys, S. H., *Theory of Probability*. third edition, Clarendon Press, Oxford, 1998.
- [19] Bernardo, J., Smith, A. F. M., *Bayesian Theory*, John Wiley & Sons, New York, 1994.
- [20] Chib, S., “Estimation and Comparison of Multiple Change-Point Models”, *Journal of Econometrics*, Vol. 86, No. 2, pp. 221-241, 1998.
- [21] Zou, C., Zhang, Y., Wang, Z., “A control chart based on a change-point model for monitoring linear profiles”, *IIE transactions*, Vol. 38, No. 12, pp. 1093-1103, 2006.
- [22] USPTO, The United States Patent and Trademark Office, <http://www.uspto.gov>, 2016, [Accessed: September 1, 2016].
- [23] WIPSON, WIPS Corporation, <http://www.wipson.com>, 2016, [Accessed: September 1, 2016].
- [24] Feinerer, I., Hornik, K., Meyer, D., “Text mining infrastructure in R”, *Journal of Statistical Software*, Vol. 25, No. 5, pp. 1-54, 2008.
- [25] Feinerer, I., Hornik, K., Package ‘tm’ Ver. 0.6, Text Mining Package, CRAN of R project, 2016.
- [26] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>, 2016, [Accessed: July 1, 2016].

## 저자 소개



### 전성해(Sunghae Jun)

1993년 : 인하대학교 통계학과 이학사

1996년 : 인하대학교 통계학과 이학석사

2001년 : 인하대학교 통계학과 이학박사

2007년 : 서강대학교 컴퓨터공학과 공학박사

2013년 : 고려대학교 정보경영공학과 공학박사

2003년~현재 : 청주대학교 통계학과 교수

관심분야 : Artificial Intelligence, Statistical Learning, Big Data

Phone : +82-43-229-8205

E-mail : shjun@cju.ac.kr