

## AI Technology Analysis using Partial Least Square Regression

JunHyeog Choi\*, Sunghae Jun\*\*

\*Professor, Dept. of Health Administration, Kimpo College, Kimpo, Korea

\*\*Professor, Dept. of Big Data and Statistics, Cheongju University, Cheongju, Korea

### [Abstract]

In this paper, we propose an artificial intelligence(AI) technology analysis using partial least square(PLS) regression model. AI technology is now affecting most areas of our society. So, it is necessary to understand this technology. To analyze the AI technology, we collect the patent documents related to AI from the patent databases in the world. We extract AI technology keywords from the patent documents by text mining techniques. In addition, we analyze the AI keyword data by PLS regression model. This regression model is based on the technique of partial least squares used in the advanced analyses such as bioinformatics, social science, and engineering. To show the performance of our proposed method, we make experiments using AI patent documents, and we illustrate how our research can be applied to real problems. This paper is applicable not only to AI technology but also to other technological fields. This also contributes to understanding other various technologies by PLS regression analysis.

▶ **Key words:** Artificial Intelligence, Partial Least Squares, Regression Analysis, Patent Data, Technology Analysis

### [요 약]

본 논문에서는 부분 최소 제곱(PLS) 회귀 모형을 이용한 인공지능(AI) 기술 분석을 제안한다. AI 기술은 이제 우리 사회의 대부분의 영역에 영향을 미치고 있다. 따라서 이 기술에 대한 정확한 이해가 필요하게 된다. AI 기술을 분석하기 위하여 전 세계 특허 데이터베이스로부터 AI 관련 특허 문서를 수집하고 텍스트 마이닝 기법을 사용하여 수집된 특허 문서에서 AI 기술 키워드를 추출한다. 본 연구에서는 추출된 AI 키워드 데이터를 PLS 회귀 모형으로 분석한다. 바이오정보학, 사회과학 및 공학 등 다양한 분야에서 고급 데이터 분석을 위하여 사용되는 PLS 회귀 모형은 부분 최소 제곱 기법을 기반으로 한다. 제안 방법의 성능을 확인하기 위하여 AI 특허 문서를 사용하여 분석 실험을 수행하고 제안하는 연구가 실제 문제에 어떻게 적용될 수 있는지 보여 준다. 본 논문은 AI 기술뿐만 아니라 다른 기술 분야에도 적용 할 수 있다.

▶ **주제어:** 인공지능, 부분최소제곱, 회귀분석, 특허데이터, 기술분석

- 
- First Author: JunHyeog Choi, Corresponding Author: Sunghae Jun
  - \*JunHyeog Choi (jhchoi@kimpo.ac.kr), Dept. of Health Administration, Kimpo College
  - \*\*Sunghae Jun (shjun@cju.ac.kr), Dept. of Big Data and Statistics, Cheongju University
  - Received: 2019. 11. 29, Revised: 2020. 02. 25, Accepted: 2020. 02. 25.

## I. Introduction

기술은 사회를 변화시키고 사회는 새로운 기술을 요구하면서 인간의 삶의 질을 향상시켜 왔다 [1]. 1차, 2차, 3차 산업혁명을 거치면서 다양한 기술이 연구 개발되어 우리 삶에 영향을 미쳐 왔다. 4차 산업혁명 이후 인공지능 (artificial intelligence, AI)는 우리 사회 대부분의 분야에 영향을 미치고 있다 [2-4]. 특정 기술에 대한 정확한 이해를 위하여 해당 기술에 대한 분석이 필요하다[1][5]. 따라서 AI 기술에 대한 이해를 위하여 AI 기술에 대한 분석이 필요하게 된다. AI는 컴퓨터가 지능을 갖도록 하는 모든 분야를 연구하는 학문으로 매우 방대한 학제적 연계를 갖는다[6]. 본 논문에서는 AI 기술 분석을 통하여 우리 사회에 많은 영향을 주고 있는 AI 기술에 대한 이해를 높이기 위한 방법에 대하여 연구한다. 기술 분석을 위하여 본 논문에서는 부분 최소제곱(partial least square, PLS) 회귀분석을 사용한다. PLS 회귀분석은 부분 최소제곱 기법을 이용하여 변수들 간의 연관성을 설명할 수 있는 모수를 추정한다[7][8]. 특히 변수들 간의 상관관계(correlation)가 존재할 경우 주 성분분석(principal component analysis, PCA)에 의한 잠재 변수(latent variable)를 이용한 변수 간 공분산(covariance) 구조에 기반 하여 회귀모형을 구축한다 [9-11]. 일반적인 회귀모형은 이와 같은 경우 다중공선성(multicollinearity) 문제 때문에 예측의 설명력이 떨어진다[12]. 본 연구에서 사용되는 AI 기술 키워드 들은 모두 AI 시스템을 구성하는 세부 기술들을 대표하기 때문에 키워드(변수) 간의 상관관계는 존재하게 된다. PLS 회귀분석을 이용한 AI 기술 분석 결과를 통하여 AI 기술에 대한 이해를 높일 수 있고 이를 바탕으로 AI 시스템을 위한 기술의 연구, 개발 과정에서 효율적인 기획이 평가가 가능하게 된다.

본 논문의 2장에서는 기존의 기술 분석 방법에 대하여 관련된 연구들을 알아본다. 제안하는 AI 기술 분석 방법은 3장에서 다루고, 4장에서는 실제 AI 기술 문서를 수집하고 분석하여 최종 결과를 도출하는 전체 과정에 대하여 설명한다. 마지막 장에서는 본 연구에 대한 결론 및 향후 연구 과제를 제시한다.

## II. Related Works

기업은 자사의 시장 경쟁력을 확보하기 위하여 끊임없이 혁신적인 기술개발에 노력한다. 국가도 마찬가지로 국가의 기술 경쟁력을 강화하기 위하여 새로운 기술의 연구

개발에 많은 국가 예산을 투입한다. 기술의 연구개발 계획을 효율적으로 세우고 평가하기 위해서는 기술에 대한 분석이 우선적으로 필요하다. 일반적으로 기술 분석은 크게 2가지 접근 방법에 의해 이루어진다. 먼저 전문가 집단 (expert group)의 경험과 지식에 의존하는 정성적 (qualitative) 방법이다. 델파이 조사(Delphi survey)가 대표적이다. 즉, 특정기술에 대하여 해당 기술 전문가 집단에게 반복적인 설문조사를 수행하고 이 결과를 요약, 정리하여 목표기술에 대한 최종 결론을 도출한다 [13][14]. 하지만 이와 같은 방법은 전문가의 주관적 판단에 의존하기 때문에 구성되는 전문가 집단에 따라 기술 분석 결과가 달라질 수 있다 [15]. 또 하나의 기술 분석 방법은 통계적 분석기법과 머신러닝 알고리즘에 의해서 이루어지는 정량적인(quantitative) 방법이다. 이 방법은 분석 대상이 되는 기술과 관련된 객관적인 데이터를 수집하고 정형화하여 통계학과 머신러닝에서 제공하는 기법을 이용하여 분석한다 [16][17]. 일반적으로 기술 분석에 사용되는 객관적인 데이터는 특허(patents)와 논문(papers)이다. 모두 텍스트에 기반한 문서 데이터이다. 따라서 텍스트 마이닝(text mining) 기법을 이용하여 수집된 특허(논문) 문서를 전처리(preprocessing)하여 정형화된(structured) 데이터를 생성하고 정량적으로 분석한다. 이것은 정성적인 방법에 비하여 상대적으로 객관적이지만 적절한 통계학 및 머신러닝 기법을 선택하고 사용할 수 있는 지식이 필요하다. 최근 R과 Python과 같은 효율적인 데이터 분석 소프트웨어가 소개되면서 정량적인 기술 분석이 활발히 수행되고 있다 [1][18][19]. 정성적 또는 정량적 기술 분석이 각각 장점과 단점을 동시에 가지고 있기 때문에 두 가지 방법을 결합한 기술 분석에 대한 연구도 이루어지고 있다 [15]. 즉, 정량적 특허분석의 결과와 전문가 집단으로부터의 조사(survey) 결과를 동시에 고려하는 기술 분석 접근 방법이다. 또 다른 기술 분석 접근 방법은 하나의 정량적 기법에 의존하지 않고 동시에 여러 가지의 정량적 기술 분석을 수행하고 그 결과들을 통합하여 결론을 도출하는 것이다. 예를 들어 자료포락분석(data envelopment analysis)과 상태공간모형(state space models)을 각각 사용한 기술 분석 결과들을 결합하여 지속가능한(sustainable) 기술 분석을 수행한 연구결과가 진행되었다 [17]. 정량적 기술 분석 모형에서는 수집된 특허 문서로부터 추출된 키워드가 변수 역할을 수행한다. 기존의 많은 기술 분석 모형에서는 각 변수(키워드)간의 독립을 가정하지만 실제로 각 키워드는 하나의 기술을 이루는 세부 요소이기 때문에 키워드 간의 상관관계가 존재한다. 따라서 키워드 간의 독립성 가정이 만족하지 않는 경우의 기술 분석 방법에 대한 연구가

필요하다. 본 논문은 이와 같은 문제를 해결하기 위하여 PLS 회귀모형을 사용한 기술 분석 방법을 제안한다.

### III. AI Technology Analysis using PLS Regression

일반적인 회귀분석은 종속변수(dependent variable)를 설명하는 독립변수들(independent variables) 간의 독립성(independence)을 중요하게 고려한다. 즉, 독립변수들 간 상관관계가 커지면 다중공선성 문제가 발생하여 모형의 설명력(예측력)이 떨어진다. 그런데 AI 기술 분석을 위하여 추출된 AI 키워드는 모두 AI 시스템을 개발하기 위하여 필요한 세부 기술을 나타내기 때문에 키워드(변수) 간의 상관성이 존재한다. 따라서 이 경우에 일반적인 선형 회귀모형을 사용하는 것은 적절하지 않다. 이와 같은 AI 키워드 데이터의 특성을 고려하여 본 논문에서는 PLS 회귀모형을 선택한다. 변수 간의 상관구조를 분석하는 PCA와 중회귀 분석(multiple linear regression)에 기반한 PLS 회귀모형은 변수 간의 상관관계가 존재하는 AI 키워드 데이터 분석을 위하여 효과적인 모형이 된다. PLS 회귀모형은 한꺼번에 많은 독립변수를 모형에 추가할 수 있기 때문에 AI 기술 키워드로 선정되는 많은 변수들을 한꺼번에 처리하는데 효율적이다. 따라서 본 논문에서는 AI 특허 키워드 데이터를 이용한 기술 분석을 위하여 PLS 회귀모형을 사용한다.

PLS 회귀분석은 독립변수 행렬  $X$ 와 종속변수 벡터  $y$  사이의 공분산 구조를 구하기 위하여 PCA에 기반한 잠재변수(latent variable)를 정의하고 이를 이용하여 다음 식과 같이  $y$ 에 대한  $X$ 의 설명력을 표현 한다 [20].

$$X = TP^t + E, \quad y = UQ^t + F \quad (1)$$

여기서  $T$ 와  $U$ 는 잠재변수를 나타내는 점수(score) 행렬이다.  $P$ 와  $Q$ 는 PCA 적재값(loadings)을 나타내고  $E$ 와  $F$ 는  $X$ 와  $y$ 의 오차(errors)를 설명하는 잔차(residuals)를 나타낸다. 이와 같은 모형 구조에서  $X$ 와  $y$ 에 대한 가중치 벡터를 구하기 위하여 주어진 데이터 표본(sample)을 이용하여 다음 식에서 정의된 표본 가중치 벡터  $w$ 와  $c$ 를 구한다 [20].

$$(\text{cov}(t, u))^2 = (\text{cov}(Xw, yc))^2 \quad (2)$$

여기서  $\text{cov}(t, u)$ 는 잠재 벡터  $t$ 와  $u$ 의 표본 공분산(sample covariance)을 나타낸다. 이 공분산 구조를 최대화하기 위하여 부분 최소제곱(partial least squares) 기법을 사용한다 [7][8]. 이와 같은 가중치 계산 결과를 이용하여  $y$ 에 대한  $X$ 의 설명력을 추정한다. 계산된 가중치를 이용하여 최종적으로 다음과 같이 개별 독립변수( $x_1, x_2, \dots, x_p$ )와 종속변수  $y$  사이의 회귀모형 식을 구한다 [7][12].

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + e \quad (3)$$

여기서  $\beta_0$ 는 절편(bias) 항이고  $\beta_i$ 는 독립변수  $x_i$ 가 종속변수  $y$ 에 미치는 영향력(가중치)를 의미하는 회귀계수이다.  $e$ 는 오차항으로 평균이 0이고 분산이  $\sigma^2$ 인 가우시안 분포(Gaussian distribution)를 따른다. 또한 개별 오차들은 각각 서로 독립이다. 본 연구에서는 개별 독립변수가 종속변수를 통계적으로 유의하게 설명하는지를 확인하기 위하여 다음과 같은 가설검정(testing)을 수행한다 [12].

$$H_0 : \beta_i = 0 \quad \text{vs.} \quad H_1 : \beta_i \neq 0 \quad (4)$$

귀무가설(null hypothesis)  $H_0$ 를 기각하지 못하면 독립변수  $x_i$ 가 종속변수  $y$ 를 통계적으로 유의하게 설명하지 못한다. 왜냐하면  $\beta_i$ 가 0이라는 가설을 기각할 수 없기 때문이다. 즉  $x_i$ 가 어떠한 값을 갖게 되어도  $\beta_i x_i$ 값은 0이 된다. 반면에  $H_0$ 를 기각하게 되면  $x_i$ 가  $y$ 를 통계적으로 유의하게 설명한다고 할 수 있다. 가설검정을 위한 검정 통계량(test statistic)  $t$ 는 다음과 같이 정의된다 [12].

$$t = \sqrt{\frac{(n-2)S_{xx}}{SSR}} \hat{\beta}_i \quad (5)$$

여기서  $n$ 은 표본 데이터의 크기이고  $S_{xx}$ 는  $x$ 의 개별값과 전체 평균과의 차이의 제곱합(sum of squares)이고,  $SSR$ 은  $y$ 의 개별값과 모형의 예측값과의 차이의 제곱합을 의미한다.  $\hat{\beta}_i$ 는  $\beta_i$ 의 추정값(estimate)이다.  $t$ 에 대한 유의확률(probability value, p-value)을 구하여 이 값이 0.1보다 작으면 90% 신뢰수준(confidence level)에서  $H_0$ 를 기각하고 해당 독립변수는 통계적으로 유의하게 종속변수  $y$ 를 설명하

게 된다. 본 논문에서 제안하는 PLS 회귀분석 기반의 AI 기술 분석은 Fig. 1과 같은 절차에 의해 수행된다.

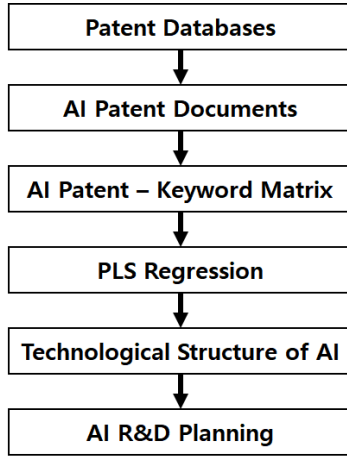


Fig. 1. AI technology analysis process

전 세계 특허 데이터베이스로부터 AI 기술 관련 특허문서를 수집부터 최종 기술 분석 결과의 활용까지의 전 과정을 나타낸다. 제안 방법의 단계별 구체적인 절차는 다음과 같다.

(Step 1) AI 기술 문서 수집 및 전처리

- (1.1) 특허 데이터베이스로부터 AI 특허 문서 수집
- (1.2) 텍스트 마이닝을 이용하여 특허문서-단어 행렬 구축
- (1.3) 특허문서-단어 행렬로부터 AI 기술 키워드 추출

(Step 2) AI 기술 키워드 분석

- (2.1) PLS 회귀분석을 이용하여 각 독립변수의 p-value 계산
- (2.2) p-value 결과를 이용하여 통계적으로 유의한 변수 선정

(Step 3) AI 기술 연관성 도출

- (3.1) PLS 회귀분석을 통하여 AI 키워드 간의 연관성 확인
- (3.2) AI 키워드에 해당하는 기술 간의 연관성을 계층화
- (3.3) 최종 AI 기술 분석 결과 도출

(Step 4) AI 기술 분석 결과의 활용

- (4.1) AI 기술에 대한 체계적 이해
- (4.2) AI 기술에 대한 연구개발 기획 및 평가에 활용

AI와 관련된 특허 문서를 수집하기 위하여 전 세계 특허 데이터베이스를 이용한다 [21][22]. 수집된 특허문서 데이터의 전처리를 통하여 정형화된 데이터인 특허문서-단

어 행렬을 구축한다. 본 연구에서는 특허문서의 전처리를 위한 텍스트 마이닝을 위하여 R 데이터 언어와 R이 제공하는 텍스트 마이닝 패키지를 사용한다 [23-25]. 특허문서-단어 행렬의 각 행은 특허문서이고 각 열은 특허문서에 1번 이상 나타나 단어이다. 이 행렬의 각 셀은 특정 단어가 각 특허문서에 나타난 빈도(frequency) 값이다. 본 논문의 PLS 회귀분석에서는 이 행렬로부터 추출된 키워드 데이터를 사용한다. 지능(intelligence)으로 시작하여 AI 시스템과 관련된 기술 키워드를 계층별로 선정하여 PLS 회귀분석을 수행한다. 단계별로 p-value를 이용하여 종속변수(키워드)에 통계적으로 유의한 영향을 미치는 독립변수(키워드)를 찾는다. 이 결과를 이용하여 다음 단계의 PLS 회귀분석을 수행한다. 단계별 PLS 회귀분석 결과를 모두 이용하여 AI 키워드 간의 기술 연관성을 찾는다. 이를 바탕으로 각 키워드가 대표하는 AI 세부기술을 정의하고 최종적으로 AI 기술 분석 결과를 도출한다. AI 기술 연관성 분석결과를 통하여 AI 기술에 대한 이해를 높이고 이를 바탕으로 AI 기술에 대한 연구개발 기획 및 평가가 효율적으로 이루어질 수 있도록 한다.

## IV. Experimental Results

본 연구에서는 전 세계의 특허 데이터베이스로부터 AI 기술 분석을 위하여 AI와 관련된 특허문서를 수집하였다 [21][22]. 최종적으로 선정된 AI 기술 유효특허문서는 13,858건이었다. R 데이터 언어와 R이 제공하는 텍스트 마이닝 라이브러리를 사용하여 특허문서-단어 행렬을 구축하였다 [23][24]. 이 행렬로부터 다음과 같이 총 38개의 AI 기술 키워드를 추출하였다: analysis, artificial, awareness, behavior, cognitive, collaborative, computing, conversation, corpus, data, dialogue, feedback, figure, image, inference, intelligence, interface, language, learning, mind, morphological, natural, network, neuro, object, ontology, pattern, recognition, representation, sentence, sentiment, situation, spatial, speech, understanding, video, vision, voice. 먼저 키워드 지능(intelligence)를 종속변수로 하고 나머지 키워드들을 독립변수로 하는 PLS 회귀분석을 수행하였다. 표 1은 분석결과와 해석을 위하여 t 통계량의 절대값(|T|)과 유의확률(p-value)을 나타낸다.

Table 1. PLS regression result: dependent=intelligence

Independent	t	p-value
analysis	1.458	0.1449
awareness	0.168	0.8662
behavior	2.151	0.0315
cognitive	0.160	0.8732
collaborative	0.067	0.9465
computing	19.322	0.0001
conversation	0.094	0.9254
corpus	0.111	0.9113
data	0.842	0.3996
dialogue	0.055	0.9558
feedback	0.481	0.6308
figure	0.005	0.9960
image	1.146	0.2519
inference	0.184	0.8543
interface	2.452	0.0142
language	0.122	0.9025
learning	4.242	0.0001
mind	0.060	0.9525
morphological	0.108	0.9142
natural	0.093	0.9256
network	1.004	0.3152
neuro	0.069	0.9446
object	1.462	0.1437
ontology	0.342	0.7324
pattern	1.261	0.2072
recognition	1.857	0.0633
representation	0.771	0.4404
sentence	0.304	0.7609
sentiment	0.073	0.9421
situation	0.196	0.8448
spatial	0.738	0.4606
speech	1.320	0.1867
understanding	0.297	0.7662
video	0.955	0.3397
vision	0.287	0.7739
voice	0.292	0.7705

본 연구에서는 90% 신뢰수준에서 각 변수(키워드) 간 통계적 유의성을 검증하였다. 즉, p-value가 0.1보다 작은 경우 해당 독립변수는 반응변수 intelligence에 유의한 영향을 미치게 된다. 표 1을 통하여 intelligence에 통계적으로 유의한 영향을 주는 AI 기술 키워드는 behavior, computing, interface, learning 그리고 recognition임을 알 수 있다. AI의 행동(behavior)과 인간과의 상호작용에 대한 점점(interface) 기술이 지능(intelligence)을 통하여 최종적으로 AI 시스템을 구축하는 데 중요한 역할을 하고 있음을 알 수 있다. 아울러 데이터(경험)의 학습(learning)을 위한 계산(computing) 및 인식(recognition) 기술을 통하여 지능 수준을 지속적으로 향상시켜 AI 시스템의 고도화에 기여할 수 있음을 확인하였다. 지능은 가장 큰 t 통계량을 갖는 computing과 가장 유의한 관계임을 알 수 있다. 표 2는 키워드 computing을 반응변수로 하는 PLS 회귀분석 결과이다.

Table 2. PLS regression result: dependent=computing

Independent	t	p-value
analysis	4.395	0.0000
awareness	0.071	0.9431
cognitive	0.411	0.6814
collaborative	0.062	0.9504
conversation	0.133	0.8943
corpus	0.151	0.8800
data	0.912	0.3615
dialogue	0.071	0.9437
feedback	0.357	0.7207
figure	0.038	0.9698
image	0.660	0.5096
inference	0.344	0.7307
language	2.565	0.0103
mind	0.066	0.9474
morphological	0.352	0.7249
natural	0.205	0.8375
network	0.465	0.6417
neuro	0.043	0.9657
object	0.027	0.9781
ontology	0.150	0.8809
pattern	0.400	0.6892
representation	2.398	0.0165
sentence	0.712	0.4765
sentiment	0.061	0.9514
situation	0.158	0.8748
spatial	0.694	0.4876
speech	0.626	0.5316
understanding	0.116	0.9077
video	0.523	0.6012
vision	0.371	0.7106
voice	0.237	0.8126

표 2에서 계산(computing) 변수에 통계적으로 유의한 키워드는 분석(analysis), 언어(language), 그리고 표현(representation)임을 알 수 있다. 데이터의 분석과 자연어 처리 기술과 분석 결과의 표현에 관한 기술을 바탕으로 한 계산 능력이 AI 시스템 구축에서 필요함을 알 수 있다. 계산 키워드와 같은 방법으로 행동(behavior), 인터페이스(interface), 학습(learning), 그리고 인식(recognition)을 종속변수로 수행한 PLS 회귀분석 결과가 표 3에 제시되었다.

Table 3. PLS regression result: dependent=(behavior, interface, learning, recognition)

Dependent	Independent
behavior	awareness, computing, image, network, object, pattern, representation, speech
interface	computing, data, image, object, ontology, speech
learning	computing, data, network, understanding
recognition	analysis, computing, object, pattern, speech, video, vision, voice

표 3의 결과에서 모든 키워드(behavior, interface, learning, recognition)에서 계산(computing)은 통계적으로 유의한 기술임을 알 수 있었다. 즉, AI 시스템 개발에서 계산 기술은 모든 세부 기술에서 필요한 분야임을 확인할 수 있다. 이미지(image) 변수는 행동과 인터페이스 변수들에 유의한 영향을 미치고, 데이터는 인터페이스와 학습에 통계적으로 유의한 영향을 끼침을 확인할 수 있다. AI 시스템에서 중요한 기술인 학습은 데이터뿐만 아니라 네트워크, 계산, 그리고 학습결과의 이해(understanding)를 지원할 수 있는 기술에 영향을 받고 있다. 또 하나의 중요한 기술인 인식(cognition)은 음성과 영상과 관련된 패턴 데이터의 분석 기술에 의존함을 알 수 있다.

본 연구에서는 일반적인 선형, 비선형 회귀분석과는 달리 PLS 회귀분석을 통하여 AI 시스템을 위한 기술 키워드 간의 기술 연관성을 확인하였다. 즉, 기존 연구들에서는 키워드 간의 독립성을 가정하였기 때문에 키워드 간의 기술 연관성을 파악하는 데 한계가 있었지만 본 연구에서는 이와 같은 문제를 해결하기 위하여 PLS 회귀분석을 이용하였다.

## V. Conclusions

본 논문에서는 최근 우리 사회의 대부분을 지배하고 있는 AI 시스템 기술에 대한 세부적인 이해를 위하여 AI 기술에 대한 정량적 분석을 수행하였다. 전 세계의 특허 데이터베이스로부터 AI 기술과 관련된 특허문서를 수집하여 텍스트 마이닝의 전처리를 통하여 특허문서-단어 행렬을 구축하고 이를 바탕으로 AI 기술 키워드를 추출하였다. AI 기술 키워드 간의 기술 연관성을 파악하기 위하여 PLS 회귀분석을 수행하였다. PLS 회귀분석 결과 AI 시스템에 직접적으로 영향을 미치는 키워드는 행동(behavior), 계산(computing), 인터페이스(interface), 학습(learning) 그리고 인식(recognition)의 5개임을 알 수 있었다. 즉 데이터의 학습과 이를 바탕으로 한 AI 시스템의 행동(서비스), 그리고 AI 시스템의 인터페이스 및 음성, 영상 등의 인식 기술이 기본적으로 AI 시스템을 구성하는 중요한 세부 기술임을 확인할 수 있었다. AI 시스템의 지속적인(sustainable) 발전을 위해서 이 키워드들과 관련된 연구 개발이 지속적으로 이루어져야 함을 알 수 있었다. 나머지 AI 키워드들과 이들 5개의 기술 키워드 간의 기술 연관성 분석 결과에 따라 계산(computing) 기술의 중요성을 확인할 수 있었고 데이터를 기반으로 하는 분석과 다양한 인식

기술의 필요성도 확인할 수 있었다. 특히 인터페이스 분야에서 객체지향 시스템 구축과 온톨로지에 대한 의존성도 확인할 수 있었다.

본 연구에서는 AI 기술 분석을 위하여 AI 관련 특허를 수집하고 분석하였다. 하지만 AI와 관련된 연구개발 결과는 논문으로도 발표되고 있다. 따라서 향후 연구에서는 특허문서뿐만 아니라 AI 관련 논문도 함께 분석하여 AI 시스템에 대한 보다 정교한 이해를 가능하게 할 것이다. 본 논문에서 사용된 PLS 회귀분석뿐만 아니라 더 다양한 기술 분석모형에 대한 연구도 함께 진행될 수 있을 것이다.

## ACKNOWLEDGEMENT

This work was supported by KIMPO College's Research Fund(2020).

## REFERENCES

- [1] A. T. Roper, S. W. Cunningham, A. L. Porter, T. W. Mason, F. A. Rossini, and J. Banks, "Forecasting and Management of Technology" Hoboken, NJ, John Wiley & Sons, 2011.
- [2] S. Y. Lee, H. M. Kim, S. H. Lee, J. H. Ha, and S. L. Lee, "AI Chatbot Providing Real-Time Public Transportation and Route Information," *Journal of The Korea Society of Computer and Information*, Vol. 24 No. 7, pp. 9-17, July 2019.
- [3] J. Jung, and J. Ahn, "Intelligent User Pattern Recognition based on Vision, Audio and Activity for Abnormal Event Detections of Single Households," *Journal of The Korea Society of Computer and Information*, Vol. 24 No. 5, pp. 59-66, May 2019.
- [4] S. Choi, T. P. Le, and T. Chung, "Controller Learning Method of Self-driving Bicycle Using State-of-the-art Deep Reinforcement Learning Algorithms," *Journal of The Korea Society of Computer and Information*, Vol. 23 No. 10, pp. 23-31, October 2018.
- [5] J. Choi, "Technology Trends for Motion Synthesis and Control of 3D Character," *Journal of The Korea Society of Computer and Information*, Vol. 24 No. 4, pp. 19-26, April 2019.
- [6] S. Russell, and P. Norvig, "Artificial Intelligence: A Modern Approach, Third Edition" Essex, UK: Pearson, 2014.
- [7] K. P. Murphy, "Machine Learning: a probabilistic perspective" Cambridge MA, MIT Press, 2012.
- [8] S. Theodoridis, "Machine Learning A Bayesian and Optimization Perspective" London UK, Elsevier, 2015.
- [9] M. G. Gustafsson, "A Probabilistic Derivation of the Partial

- Least-Squares Algorithm,” *Journal of Chemical Information and Computer Sciences*, 41(2), 288-294, February 2001.
- [10] L. Sun, S. Ji, S. Yu, and J. Ye, “On the Equivalence Between Canonical Correlation Analysis and Orthonormalized Partial Least Squares,” *Proceedings of the 21st international joint conference on Artificial intelligence*, pp. 1230-1235, 2009.
- [11] M. Barker and W. Rayens, “Partial least squares for discrimination,” *Journal of Chemometrics*, Vol. 17, No. 3, pp. 166-173, March 2003.
- [12] S. M. Ross, “*Introductory Statistics, Fourth Edition*” London, UK, Academic Press Elsevier, 2017.
- [13] J. Keller, and H. A. V. D. Gracht, “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, June 2014.
- [14] H. A. Linstone, and M. Turoff, “Delphi: A Brief Look Backward and Forward,” *Technological Forecasting and Social Change*, Vol. 78, Iss. 9, pp. 1712-1719, November 2011.
- [15] S. Jun, S. J. Lee, J. B. Ryu, and S. Park, “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, October 2015.
- [16] S. Park, and S. Jun, “Technology Analysis of Global Smart Light Emitting Diode (LED) Development Using Patent Data,” *Sustainability*, Vol. 9, Iss. 8, pp. 1363, August 2017.
- [17] J. Kim, B. Sun, and S. Jun, “Sustainable Technology Analysis Using Data Envelopment Analysis and State Space Models,” *Sustainability*, Vol. 11, Iss. 13, pp. 3597, June 2019.
- [18] J. Kim, N. Kim, Y. Jung, and S. Jun, “Patent data analysis using functional count data model,” *Soft Computing*, Vol. 23, Iss. 18, pp. 8815-8826, September 2019.
- [19] S. Jun, “Bayesian Count Data Modeling for Finding Technological Sustainability,” *Sustainability*, Vol. 10, No. 9, pp. 3220, September 2018.
- [20] R. Rosipal, and N. Krámer, “Overview and recent advances in partial least squares,” *Proceedings of International Statistical and Optimization Perspectives Workshop*, pp. 34-51, 2005.
- [21] USPTO, The United States Patent and Trademark Office, <http://www.uspto.gov>, 2019.
- [22] WIPSON, WIPS Corporation. <http://www.wipson.com>, <http://global.wipscorp.com>, 2019.
- [23] R Development Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org>, 2019.
- [24] I. Feinerer, K. Hornik, and D. Meyer, “Text mining infrastructure in R,” *Journal of Statistical Software*, Vol. 25, No. 5, pp. 1-54, March 2008.
- [25] I. Feinerer, and K. Hornik, Package ‘tm’ Ver. 0.7-6, Text Mining Package, CRAN of R project, 2019.

## Authors



JunHyeog Choi received B.S. degrees from Department of Computer Science, Kyonggi University, He also received M.S., PhD degree from Department of Computer Science & Engineering, Inha University, Korea in 1995,

2000, and M.S., PhD from Graduate School of Management of Technology, Yonsei University, Seoul, Korea in 2003, 2013. He is Professor in the Department of Health Administration, Gimpo, Gyeonggi-do, Korea. His current research interests include Big data analysis and data science.



Sunghae Jun received B.S., M.S., and PhD degrees from Department of Statistics, Inha University, Incheon, Korea in 1993, 1996, and 2001, respectively. He also received PhD degree from Department of Computer Science,

Sogang University, Seoul, Korea in 2007, and PhD from Information Management Engineering from Korea University, Seoul, Korea in 2013. He is Professor in the Department of Statistics, Cheongju University, Chungbuk, Korea. Also, He was visiting scholar in Department of Statistics, Oklahoma State University, Stillwater, Oklahoma, USA from 2009 to 2010. His current research interests include AI and data science.