

# DTW-kNN 기반의 유망 기술 식별을 위한 의사결정 지원 시스템 구현 방안

정도현<sup>1</sup>, 박주연<sup>2\*</sup>

<sup>1</sup>덕성여자대학교 글로벌융합대학 교수, <sup>2</sup>덕성여자대학교 차미리사교양대학 교수

## Implementation of DTW-kNN-based Decision Support System for Discriminating Emerging Technologies

Do-Heon Jeong<sup>1</sup>, Ju-Yeon Park<sup>2\*</sup>

<sup>1</sup>Professor, College of Global Convergence Studies, Duksung Women's University

<sup>2</sup>Professor, College of Cha Mirisa Liberal Arts, Duksung Women's University

**요약** 본 연구는 기계 학습 기반의 자동 분류 기법을 적용함으로써 유망 기술의 선정 과정에 활용할 수 있는 의사결정 지원 시스템의 구현 방안을 제시하는 것을 목표로 한다. 연구 수행을 위해 전체 시스템의 아키텍처를 구축하고 세부 연구 단계를 진행하였다. 우선, 유망 기술 후보 아이템을 선정하고 빅데이터 시스템을 활용하여 추세 데이터를 자동 생성하였다. 기술 발전의 개념 모델과 패턴 분류 체계를 정의한 후 자동 분류 실험을 통해 효율적인 기계 학습 방안을 제시하였다. 마지막으로 시스템의 분석 결과를 해석하고 활용 방안을 도출하고자 하였다. 본 연구에서 제안한 동적 시간 와핑(DTW) 기법과 k-최근접 이웃(kNN) 분류 모델을 결합한 DTW-kNN 기반의 분류 실험에서 최대 87.7%의 식별 성능을 보여주었으며, 특히 추세의 변동이 심한 'eventual' 정의 구간에서는 유클리디언 거리(ED) 알고리즘 대비 39.4% 포인트의 최대 성능 차이를 보여주어 제안 모델의 우수함을 확인할 수 있었다. 또한, 시스템이 제시하는 분석 결과를 통해, 대량의 추세 데이터를 입력받아 유형별로 자동 분류하고 필터링하는 과정에 본 의사결정 지원 시스템을 효과적으로 활용할 수 있음을 확인하였다.

**키워드** : 유망 기술, 동적 시간 와핑, 기계 학습, 자동 분류, 의사결정 지원 시스템

**Abstract** This study aims to present a method for implementing a decision support system that can be used for selecting emerging technologies by applying a machine learning-based automatic classification technique. To conduct the research, the architecture of the entire system was built and detailed research steps were conducted. First, emerging technology candidate items were selected and trend data was automatically generated using a big data system. After defining the conceptual model and pattern classification structure of technological development, an efficient machine learning method was presented through an automatic classification experiment. Finally, the analysis results of the system were interpreted and methods for utilization were derived. In a DTW-kNN-based classification experiment that combines the Dynamic Time Warping(DTW) method and the k-Nearest Neighbors(kNN) classification model proposed in this study, the identification performance was up to 87.7%, and particularly in the 'eventual' section where the trend highly fluctuates, the maximum performance difference was 39.4% points compared to the Euclidean Distance(ED) algorithm. In addition, through the analysis results presented by the system, it was confirmed that this decision support system can be effectively utilized in the process of automatically classifying and filtering by type with a large amount of trend data.

**Key Words** : Emerging Technology, Dynamic Time Warping, Machine Learning, Automatic Classification, Decision Support System

This Research was supported by Duksung Women's University Research Grants 2022 (300007084).

\*Corresponding Author : Ju-Yeon Park(juyeonpark@duksung.ac.kr)

Received July 7, 2022

Revised August 4, 2022

Accepted August 20, 2022

Published August 28, 2022

## 1. 서론

기술 중심의 사회가 도래하면서 많은 연구자들이 산업 및 연구 분야에서 세상을 바꾸는 혁신적 기술 발굴에 박차를 가하고 있다. 이러한 끊임없는 기술 기회 발굴(technology opportunity discovery; TOD)을 통해 향후 시장을 점유할 수 있는 수준의 기술, 즉 와해성 기술(disruptive technology) 발굴에 학계를 포함한 각국의 기업과 정부가 적극적으로 가세하고 있다[1]. 메사추세츠 공과대학(Massachusetts Institute of Technology; MIT)에서는 MIT Technology Review를 통해 매년 10대 혁신 기술(10 Breakthrough Technologies)을 선정 발표하고 있다[2]. 세계적인 시장 조사 분석 전문기업인 가트너(Gartner)는 독자적인 하이프 사이클(Hype Cycle) 모형을 통해 매년 기술 성장 수준을 진단하고 예측한다[3]. 우리 정부 역시 유망 핵심 기술의 R&D 발전 전략을 해마다 발표하며 유망 기술의 선정과 연구 개발에 박차를 가하고 있다[4]. 이러한 각 계의 노력에도 불구하고, 기술 분야의 발전에 따른 핵심 요소 기술의 폭발적 증가 현상으로 인해 유망 기술 아이টে임을 발굴하는 의사결정의 어려움 역시 크게 증가하고 있다[5]. 빅데이터 기반의 유망 기술 진단 시스템 구축과 활용의 필요성이 커지고 있으나 관련 연구는 충분히 수행되고 있지 않은 실정이다.

이에 본 연구는 유망 기술의 선정 과정에 활용할 수 있는 의사결정 지원 시스템의 구현 방안을 제안하는 것을 최종 목표로 한다. 본 연구는 의의는 다음과 같다. 첫째, 시계열 패턴 데이터의 유사도를 측정하는 효율적인 기법을 제안하고자 한다. 이를 위해 다양한 추세 패턴을 정의한 분류 체계를 구축한 후, 성능 평가 실험을 통해 기존의 모델 대비 제안 모델의 우수성을 검증하고자 한다. 둘째, 유망 기술을 식별하는 의사결정 지원 시스템의 아키텍처를 제안함으로써, 대량의 추세 데이터를 자동 분류하고 필터링하는 과정에 본 시스템을 효과적으로 활용할 수 있음을 제시하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 이론과 연구 사례를 소개하고 3장에서는 유망 기술 식별을 위한 의사결정 지원 시스템의 아키텍처를 제안하고 수행 과정을 설명한다. 4장에서는 실험을 통해 시스템의 식별 성능을 측정하고, 제안 시스템의 활용 방안을 제시한다. 마지막으로 연구의 수행 결과와 향후 연구 계획을 언급한다.

## 2. 관련 연구

### 2.1 유망 기술 성숙도 진단 모델

유망 기술들의 성숙도 진단을 위한 대표적인 기술 발전 모형으로 하이프 사이클과 Great Surges of Development(이하 GSD)를 들 수 있다. 가트너가 제안한 하이프 사이클은 자사의 정보 시스템을 통해 연구 이슈 아이টে임을 발굴하고 기술 기회를 분석하기 위해 활용되어 왔다[6, 7]. Fig. 1과 같이 기술 발전 정도를 총 5단계로 구분하여, 유망 기술이 태동하고(innovation trigger) 기대감이 최대치에 오른 후(peak of inflated expectations) 기술 한계를 극복하고 재정지원을 얻은 기술이(slope of enlightenment) 안정적인 생산 수준에 이르는(plateau of productivity) 과정으로 설명한다.

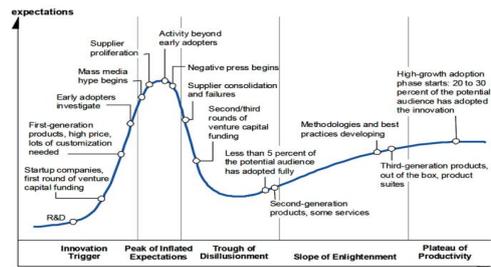


Fig. 1. The hype cycle and its stage indicators[6, 7]

하이프 사이클 모형에 비해 저작권이 자유로운 GSD 모형은 금융 분야의 거래 추이 분석[8], 기술의 거시적 흐름과 변화 분석[9] 등 다양한 연구에 활용되고 있다. Fig. 2와 같이 GSD 역시 발전 단계를 irruption, frenzy, turning point, synergy, maturity의 5단계로 정의하고 있다[10]. 크게 installation과 deployment 시기로 구분하고 turning point 단계에서 재정지원과 기술적 잠재력을 바탕으로 synergy 단계를 거쳐 안정 단계(maturity)에 이른다고 설명한다. 유망 기술의 발전 과정에서 두 모델이 공통적으로 주목하는 부분은 기술의 초기 발전세가 둔화되고 상용화 수준의 기술력과 재정 확보가 필요한 3단계(trough of disillusionment와 turning point)이며 유망 기술 여부 판단의 중요한 지점이 된다.

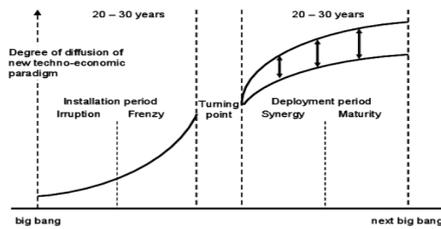


Fig. 2. Five steps of "Great Surges of Development" [10]

## 2.2 DTW 유사도 측정 기법과 kNN 분류 기법

시계열 데이터 간의 거리 측도로 많이 사용되는 유클리디언 거리(euclidean distance; 이하 ED) 알고리즘은 시간의 축이 뒤틀어진 상황에서는 유사도의 계산 정확도가 매우 낮아진다. 반면, 비선형 방식의 동적 시간 와핑(dynamic time warping; 이하 DTW) 알고리즘은 이러한 시간 축의 비틀림을 교정하여 시간의 기준점이 다르거나 시간 축의 스케일 차이로 성능이 저하되는 단점을 보완할 수 있다. DTW 기법의 주요 알고리즘은 다음과 같다. 길이가  $m, n$  인 두 시계열 데이터  $X, Y$ 에 대해,  $m \times n$  행렬을 생성하고 거리  $|x_i - y_j|$ 를 측정한다. 경계 조건, 단조 증가성, 연속성의 조건을 만족하며,  $\min()$ 함수의 값이 최소가 되도록 하는 와핑 경로  $w$ 를 찾아낸다. 이 때 공식 (1)의  $k$ 번째 와핑 경로인  $w_k$ 의 누적 와핑 거리 식  $D(i, j)$ 는 최단 경로를 결정하는 최종 유사도를 의미하며 공식 (2)와 같이 정의된다[11].

$$DTW(X, Y) = \frac{1}{K} \sqrt{\sum_{k=1}^K w_k} \quad (1)$$

$$D(i, j) = d(x_i, y_j) + \min \left\{ \begin{array}{l} D(i-1, j-1) \\ D(i-1, j) \\ D(i, j-1) \end{array} \right\} \quad (2)$$

본 연구는 효율적인 유사도 측정 기법인 DTW와 함께 자동 분류 기법으로 k-최근접 이웃(k-nearest neighbors; 이하 kNN) 알고리즘을 활용하고자 한다. kNN 기법은 지도 학습(supervised learning) 기법 중 하나로 분류 대상 데이터와 가장 가까운 k개의 데이터를 참고하여 가장 많이 투표된 범주를 최종 선택한다. kNN 기법은 알고리즘이 간단하고 학습 모델을 생성하지 않으므로 속성들의 관계를 고려하지 않는다는 특징이 존재한다. 이러한 특징으로 인해 데이터 마이닝 분야

에서 DTW 기법과 함께 시계열 데이터의 분류 성능 향상을 위한 알고리즘 연구[12, 13]에 활용되며, 검색 키워드의 이벤트 탐지 및 분석 등의 응용 연구[14]에 활용되기도 한다.

## 2.3 기계 학습 기반 유망 기술 발굴 연구

기계 학습 기반의 유망 기술 발굴 연구는 전문가 중심의 선정 방식에 비해 대량의 데이터를 빠르고 일관성 있게 처리할 수 있다는 장점이 있어 많은 연구가 진행되어 왔다. 주로 전문가의 의사결정 과정을 지원하기 위한 용도로 활용되며, 방법론에 따라 범주 체계가 정의되지 않은 상태에서 군집화(clustering)를 수행하는 비지도 또는 반지도 학습(un-/ semi-supervised learning) 유형과 본 연구와 같이 범주화(categorization)된 결과를 제공하는 지도 학습(supervised learning) 유형으로 구분될 수 있다.

기술 기회 발굴을 위해 용어 가중치 모델인 TF-IDF (term frequency-inverse document frequency)를 기반으로 k-means 클러스터링 기법을 적용한 비지도 학습 연구[15]와 딥러닝(deep learning)의 doc2vec 기법을 이용한 반지도 학습 기반의 지식 그래프 활용 연구[16] 등이 수행된 바 있다. 용어 지배 값(term dominance value) 기법과 패턴 양자화(quantization)를 통해 추세 데이터를 자동 분류하는 지도 학습 기반 연구가 수행되기도 하였다[17].

이와 같이 유망 기술 발굴과 관련하여 다양한 기법의 연구가 수행되고 있는 반면, 시스템의 구축 및 활용 방안을 구체적으로 제시한 연구는 거의 수행된 바가 없다. 따라서 본 연구에서는 효율적 기계 학습 방안의 제안과 더불어 통합 시스템의 개발 및 활용 방안에 대해 논의하고자 한다. 다음 장에서는 본 연구에서 제안하는 의사결정 지원 시스템의 아키텍처와 연구 수행 과정을 상술하고자 한다.

## 3. 유망 기술 식별을 위한 의사결정 지원 시스템

### 3.1 시스템 구성 및 연구 절차

본 연구의 목적은 기술들의 발전 추세를 효율적으로 분석하기 위한 방안과 시스템을 제안함으로써, 연구자는 유망 연구 아이템을 발견하고 민간기업과 공공 및 정부 기관은 유망 기술 발전 전략을 수립할 수 있도록 지원하

는 것이다. 이를 위해 Fig. 3과 같이 기계 학습 기반의 의사결정 지원 시스템의 아키텍처를 구성하고 이 절차에 따라 연구를 수행 하였다. 본 시스템은 크게 3단계로 구성된다. 첫 번째는 데이터 수집 단계(collect)이며, 실시간 또는 배치 방식을 통해 다양한 외부 자원으로부터 대량의 기술명 정보를 수집, 가공한다. 두 번째는 데이터 처리 단계(data processing)이다. Open API 기반의 외부 빅데이터 시스템을 활용해 추세 정보를 생성하는 데이터 퓨전(data fusion) 과정과 기술 발전 패턴의 정의 모형을 작성하는 과정으로 구성된다. 세 번째는 기계 학습을 통해 데이터 분석 결과를 제공하는 단계(analytics)이다. 실험을 통해 시계열 패턴 간의 유사도를 측정하는 효율적 기법을 선정하고, 자동 분류 결과와 시각화 데이터를 제공함으로써 의사결정 지원을 수행할 수 있도록 한다. 단계별로 수행된 세부 내용을 이후 상세히 소개하고자 한다.

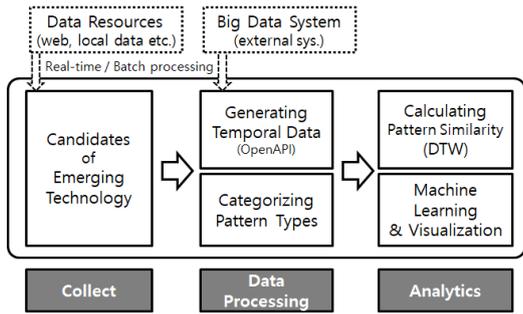


Fig. 3. Architecture of the decision support system

3.2 데이터 수집 및 구축

유망 기술의 패턴 식별 실험을 위한 데이터 수집을 위해 Wikipedia의 List of Emerging Technologies 페이지([https://en.wikipedia.org/wiki/List\\_of\\_emerging\\_technologies](https://en.wikipedia.org/wiki/List_of_emerging_technologies))를 활용하였다. 목록에는 Agriculture 분야의 agricultural robotics 기술부터 Transport 분야의 vehicular communication systems 기술까지 총 15개의 기술 분야에 걸쳐 약 200여 개의 유망 기술 정보가 요약되어 있다. 또한 Status 항목에 현 발전 수준을 hypothetical, experiments, prototypes, (early) research, development, trial projects, (early) commercialization 등으로 상술하고 있어 기술 성숙도 판단 시 참고 자료로 유용하다.

본 연구를 위해 15개 전 분야에 걸쳐 유사 동의어를

포함한 모든 유망 기술 용어를 수집하여 총 205개의 기술명 목록을 구축하였다. 작성된 목록을 이용해 각 기술명의 학술 논문의 연차별 게재 정보를 자동 수집하기 위해 과학기술 학술 정보를 제공하는 Open API 기반의 ScienceOn 서비스(<http://scienceon.kisti.re.kr>)를 활용하였다. 데이터 수집과 처리의 모든 과정은 파이썬 프로그래밍 언어로 직접 작성하였다.

2007-2021년 기간 동안 발생한 논문 건수의 기술 통계는 Table 1과 같다. 수집 대상 기술은 총 205종이며, 중앙값은 5,230.4, 표준편차는 17,941.4로 유망 기술의 특성상, 관련 연구 분야와 기술의 활성 정도에 따라 편차가 크게 나타나는 특징이 나타나고 있다.

Table 1. Descriptive Statistics for Temporal Data of Emerging Technologies (from 2007 to 2021)

	value		value
total cnt.	205	Q1 (25%)	20
mean	5,230.4	Q2 (50%)	220
std	17,941.4	Q3 (75%)	1,780
min.	1	max.	158,357

3.3 유망 기술 개념 모델과 패턴 정의

유망 기술을 진단하고 선정하기 위한 기계 학습 환경을 구축하기 위해, 우선 Fig. 4와 같이 기술의 발전 단계를 개념적으로 정의하는 과정이 필요하였다. 우선, 사분위수(quartiles)를 바탕으로, 기술의 개념이 태동하여 발전을 시작하는 단계인 하위 25%의 1사분위(Q1) 구간은 2007-2021년 기간 내 총 20건 이하의 저빈도 데이터이므로 본 연구의 추세 식별 실험에서 제외하였다. 이 구간은 미래유망 기술로 간주되나 연구가 본격화되기 전인 가설 단계의 기술(hypothetical technologies)로 정의할 수 있다[18]. 1사분위에서 3사분위 사이의 중간 빈도 기술군은 발전 과정에 있는 상태이며, 3사분위 이상의 성숙 단계 기술 중 일부는 매우 높은 발생 빈도와 지속적으로 높은 성숙도를 보이는데, 이러한 유망 기술들은 이후 시장에서 주류가 되는 완성 수준의 기술(matured technologies)이 되며, 유망 기술의 판단 범주에서 벗어나게 된다. 따라서 본 연구에서는 Q1에서 Max까지를 유망 기술 진단 실험의 대상으로 정의하였다.

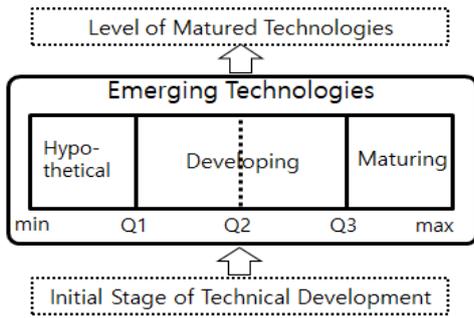


Fig. 4. Conceptual model for the level of technological development based on quartiles

구축한 개념 모델을 바탕으로 자동 분류 실험을 위한 학습데이터 구축을 위해 Table 2와 같이 패턴 유형을 정의하였다. 상위 그룹(Group)은 성숙과정( $G1$ ), 지속성장( $G2$ ), 발전가능( $G3$ ), 판단보류( $G4$ )의 네 단계로 크게 구분하고, 패턴 유형에 따라 7개의 하위 범주(class)로 정의하였다. 특히 C4와 C5는 시간의 흐름에 따라 발전 형태로 변화가 가능한 잠재적인 유망 기술 후보 그룹으로 간주할 수 있다. 앞서 언급한 바와 같이 최저빈도 그룹인 1사분위 구간의 기술 51종은 가설 단계의 기술(hypothetical)로 정의하고 실험에서 제외하였다.

Table 2. Pattern definition for automatic classification

Group	Class	Trend pattern	Type
G1	C1		maturing
G2	C2		increasing
	C3		rising
G3	C4		reoccurring
	C5		eventual
G4	C6		irregular
	C7		decreasing
N/A			hypothetical

## 4. 시스템 성능 평가 및 분석

### 4.1 실험 및 개발 환경

수집된 시계열 데이터는 기계 학습을 위한 입력용 데이터 프레임(data frame)으로 변환 생성하였다. 데이터 프레임 생성에는 파이썬의 판다스(Pandas) 라이브러리를 활용하였으며, 시각화 차트를 자동 생성하기 위해 matplotlib 라이브러리를 사용하였다. 복잡하고 다양한 형태로 나타나는 유망 기술의 시계열 패턴 분석에 DTW 기법이 효과적임을 확인하기 위해, ED를 베이스라인으로 선정하여 성능을 비교 측정하고자 하였다. DTW 알고리즘의 지수적 연산( $O(n^2)$ )을 선형적 연산( $O(n)$ )으로 변환하여 처리 성능을 개선한 FastDTW 모델[19]을 활용하였고, 파이썬의 numpy, scipy와 fastdtw 등 다양한 라이브러리를 사용하여 분석 도구를 직접 개발하였다.

### 4.2 분류 모델 및 평가 방법

본 연구에서는 다수의 데이터 마이닝 연구에서 DTW 기법과 함께 사용하는 분류기인 kNN 분류 모델을 활용하였다. kNN 모델의 성능 비교 연구를 통해 1NN 모델이 높은 식별 성능을 보였던 기존의 연구 사례[20]를 바탕으로, 본 연구에서도  $k(\text{neighbors})=1$ , similarity measure = dtw, euclidean distance를 주요 파라미터로 사용하였다.

실험을 통해 7가지 클래스(class)의 패턴 식별 성능과 성숙과정, 지속성장, 발전가능, 판단보류로 정의되는 4가지 상위 그룹(group)의 패턴 식별 성능을 각각 측정하였다. 우선 데이터의 관측과 예측 결과를 이용하여 Table 3, 공식 (3)과 같이 kNN 분류기가 긍정으로 예측한 전체 값( $TP + FP$ ) 중 긍정으로 올바르게 예측한 값( $TP$ )의 비율인 정확률(accuracy)을 산출하였다. 또한 두 가지 측정 방식인, 전체 개체에 대한 마이크로 평균(micro-averaged) 정확률 및 범주별 정확률의 평균값인 매크로 평균(macro-averaged) 정확률을 모두 산출하였다.

평가를 위한 교차 검증(cross validation) 방법은 k-fold 검증을 사용하였으며 비교적 작은 규모의 데이터 평가에 적합한 leave-one-out 교차 검증(LOOCV)을 사용하였다. 이 방법은 식별 대상 1건을 제외한 나머지를 학습 셋으로 활용하므로 반복 실험의 횟수( $n$ )는 많으나 안정적인 평가 결과를 얻을 수 있는 장점이 있다.

Table 3. Confusion matrix

		Observation	
		YES	NO
Prediction	YES	TP (true positive)	FP (false positive)
	NO	FN (false negative)	TN (true negative)

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

### 4.3 시스템 성능 평가

베이스라인 알고리즘인 ED와 본 연구에서 제안하는 DTW 기법을 비교한 전체 실험 결과는 Table 4, Fig. 5와 같다. Table 4에는 패턴별로 구분한 7개의 클래스별 성능을 표기하였으며, 이를 바탕으로 클래스 정확률을 평균한 매크로 평균 정확률(Mac)을 계산하였다. Fig. 5는 클래스와 그룹을 구분하여 ED와 DTW의 성능을 전체적으로 비교한 것이다.

Table 4. Comprehensive performance comparisons (P: precision, Mac: macro-averaged, Mic: micro-averaged)

Class	ED	DTW	Group	ED	DTW
	P(%)	P(%)		P(%)	P(%)
C1	75.0	100.0	G1	75.0	100.0
C2	76.1	76.9	G2	85.3	91.2
C3	75.0	83.9	G3	55.0	81.3
C4	42.9	80.0			
C5	33.3	72.7	G4	75.0	68.4
C6	62.5	66.7			
C7	75.0	70.0			
Mac	62.8	78.6	Mac	72.6	85.2
Mic	69.5	79.2	Mic	79.2	87.7

실험 결과, DTW 기법을 적용한 4개 그룹에 대한 마이크로 평균 정확률 결과가 87.7%로 최고치를 보여주었다. DTW 알고리즘을 사용한 식별 결과는 7개 클래스 기준의 분류 체계에서 매크로 평균 정확률(Mac)과 마이크로 평균 정확률(Mic)에서 각각 15.8% 포인트와 9.7% 포인트의 성능 차이를 보였다. 또한 4개 그룹 기준의 분류 체계에서도 12.3% 포인트(Mac 항목), 8.7% 포인트(Mic 항목)의 차이를 보여 모두 ED 알고리즘에 비해 전체적으로 우수한 결과가 나타났다. 특히, 시간 축의 변화가 심한 C4, C5(G3 구간)에서 각각 37.1, 39.4% 포인트의 큰 성능 차이가 발생한 것으로 나타나고 있어 DTW의 동적 알고리즘에 기반한 패턴 처리 특성이 우수한 결과로 나타났음을 알 수 있다.

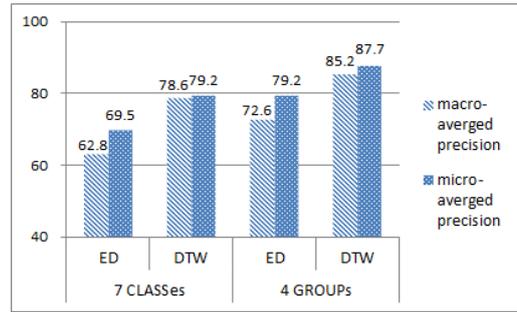


Fig. 5. Overall performance comparison btw. ED & DTW

### 4.4 시스템 분석 결과 및 활용 방안

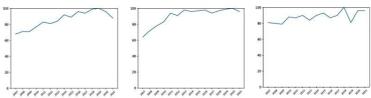
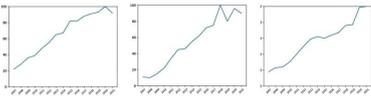
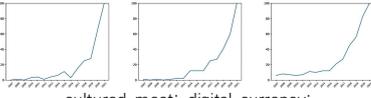
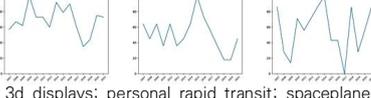
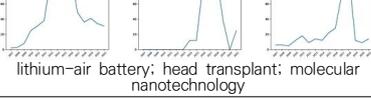
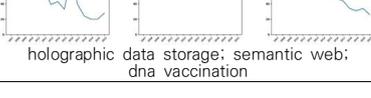
Table 5는 시스템을 통해 자동 생성되어 추천된 클래스별 대표 기술이다. 성숙과정(G1)에 접어든 C1 범주에 해당하는 대표적인 기술로는 fullerene, carbon nanotubes, tissue engineering이 나타났으며, 지속성장(G2) 그룹에 해당하는 C2와 C3에서는 각각 magnetic nanoparticles, nanomedicine, robotic surgery와 cultured meat, digital currency, immersive virtual reality가 대표적인 기술로 나타나고 있다. 다음은 성장가능(G3) 그룹에 해당하는 C4와 C5 유형이다. 현재는 뚜렷한 패턴이 보이지 않으나 시간이 지남에 따라 성장 추세를 보일 수 있는 후보 기술군으로 3d displays, personal rapid transit, spaceplane 과 lithium-air battery, head transplant, molecular nanotechnology 등이 대표적인 기술로 나타났다. 마지막 판단보류(G4) 그룹과 관련하여, 패턴이 잘 드러나지 않는 C6의 plantibody, fusion rocket, mass driver나 유망 기술 후보임에도 실재로는 하락세가 보이는 C7의 기술들(holographic data storage, semantic web, dna vaccination)은 필터링 대상 기술로 구분된다.

이와 같이, 본 시스템은 기계 학습 기법을 기반으로 유망 기술을 진단하는 데 효율적이며, 의사결정 과정에서 불필요한 데이터를 사전에 자동 필터링할 수 있는 장점이 있다. 또한, Table 5와 같이 패턴의 유형 별로 대량의 기술들을 자동 분류할 수 있으며, 특정한 기술을 질의로 입력하여 가장 유사한 패턴을 보이는 특정 기술을 식별하여 제공할 수도 있다. 분석의 목적에 따라 시스템이 제공하는 식별 결과를 다양하게 해석하고 활용할 수 있을 것으로 기대한다.

시스템 운용 시 발생할 수 있는 사항으로, 외부 정보

자원을 추가하여 상호 참조하거나 시스템에 도입하는 방안이 필요할 수 있다. 전문 분야별로 존재하는 요소 기술 간의 계층 정보인 기술 트리(tech. tree)를 활용하거나, 논문 외에 연구비 지원 정보, 특허 정보 등을 추가 도입하여 자원의 다양성을 확보할 수 있을 것이다. 향후 외부 자원을 본 시스템과 융합하는 방안에 대한 추가적인 연구가 필요할 것으로 보인다.

Table 5. Representatives by class (3 examples per class)

Group / Class	Real patterns of technology
G1 C1	 fullerene; carbon nanotubes; tissue engineering
G2 C2	 magnetic nanoparticles; nanomedicine; robotic surgery
C3	 cultured meat; digital currency; immersive virtual reality
G3 C4	 3d displays; personal rapid transit; spaceplane
C5	 lithium-air battery; head transplant; molecular nanotechnology
G4 C6	 plantibody; fusion rocket; mass driver
C7	 holographic data storage; semantic web; dna vaccination

## 5. 결론

본 연구는 빅데이터 환경에서 대량으로 유입되는 유망 기술 후보들의 기술 성숙도를 효율적으로 자동 진단

할 수 있는 기계 학습 기반의 의사결정 지원 시스템의 구현 방안을 제안하는 것을 목표로 한다.

연구 수행을 위해 우선 전체 시스템의 아키텍처를 구축하고 세부 연구 단계를 진행하였다. 첫째, 유망 기술 후보 아이템을 선정하고 Open API 시스템을 활용하여 추세 데이터를 자동 생성하였다. 또한, 식별 실험을 위해 개념 모델을 작성하고 다양한 기술 발전의 패턴을 정의하였다. 둘째, 효율적인 기계 학습 방안을 제시하였다. 본 연구에서 제안하는 DTW-kNN 모델 기반의 분류 실험에서 최대 87.7%의 식별 성능을 보여주었다. 특히, 추세의 변동이 심한 'eventual(C5)' 구간에서는 베이스라인인 ED 알고리즘 대비 39.4% 포인트의 성능 차이를 보여주어 제안 모델의 우수함을 확인할 수 있었다. 마지막으로, 시스템이 제시하는 분석 결과를 통해, 대량의 추세 데이터를 입력받아 자동으로 유형별로 분류하고 필터링하는 과정에 본 시스템을 효과적으로 활용할 수 있음을 확인하였다.

향후 연구에서는 기계 학습 알고리즘의 개선, 다양한 외부 자원의 활용 등을 통해 식별 성능을 고도화하는 방안을 마련하고자 한다. 또한 특정 분야의 기술 트리를 심층 분석하는 실용적 연구를 수행할 예정이다.

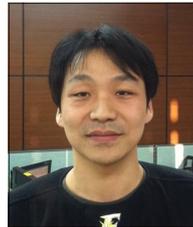
## REFERENCES

- [1] I. Park & B. Yoon. (2018). Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *Journal of Informetrics*, 12(4), 1199-1222. DOI : 10.1016/j.joi.2018.09.007
- [2] MIT. (2022). *2022 10 Breakthrough Technologies*. MIT Technology Review. <https://www.technologyreview.com/2022/02/23/1045416/10-breakthrough-technologies-2022/>
- [3] D. E. O'Leary. (2008). Gartner's Hype Cycle and Information System Research Issues. *International Journal of Accounting Information Systems* 9(4), 240-252. DOI : 10.1016/j.accinf.2008.09.001
- [4] KISTEP. (2022). *A Study on the Selection of KISTEP Future Key Technologies in 2022*. KISTEP. [https://www.kistep.re.kr/reportDetail.es?mid=a10305020000&rpt\\_tp=831-006&rpt\\_no=RES0220220075](https://www.kistep.re.kr/reportDetail.es?mid=a10305020000&rpt_tp=831-006&rpt_no=RES0220220075)
- [5] D. H. Jeong & H. S. Joo. (2018). Discovering Interdisciplinary Convergence Technologies Using

- Content Analysis Technique Based on Topic Modeling. *Journal of the Korean Society for information Management*, 35, 77-100. DOI : 10.3743/KOSIM.2018.35.3.077
- [6] J. Fenn & M. Raskino. (2009). *Understanding Gartner's Hype Cycles*. Gartner. <https://www.gartner.com/en/documents/1069314>
- [7] O. Dedehayir & M. Steinert. (2016). The hype cycle model: A review and future directions. *Technological Forecasting and Social Change*, 108, 28-41. DOI : 10.1016/j.techfore.2016.04.005
- [8] A. Caiani, A. Godin & S. Lucarelli. (2015). *Innovation and Finance: A Stock Flow Consistent Analysis of Great Surges of Development*. In The Evolution of Economic and Innovation Systems. Springer. DOI : 10.1007/978-3-319-13299-0\_17
- [9] L. Kanger & J. Schota. (2019). Deep transitions: Theorizing the long-term patterns of socio-technical change. *Environmental Innovation and Societal Transitions*, 32, 7-21. DOI : 10.1016/j.eist.2018.07.006
- [10] C. Perez. (2003). *Technological Revolutions and Financial Capital: The Dynamics of Bubbles and Golden Ages*. Cheltenham, UK: Cambridge University Press. DOI : 10.1017/S002205070348193X
- [11] E. Keogh. (2005). Exact indexing of dynamic time warping. *Knowledge and Information Systems*, 7(3), 358-386. DOI : 10.1007/s10115-004-0154-9
- [12] K. Yang & C. Shahabi. (2007). An efficient k nearest neighbor search for multivariate time series. *Information and Computation*, 205(1), 65-98. DOI : 10.1016/j.ic.2006.08.004
- [13] H. H. Hsu, A. C. Yang & M. D. Lu. (2011). KNN-DTW Based Missing Value Imputation for Microarray Time Series Data. *Journal of Computers*, 6(3), 418-425. DOI : 10.4304/jcp.6.3.418-425
- [14] D. H. Jeong. (2019). Enhancing Classification Performance of Temporal Keyword Data by Using Moving Average-based Dynamic Time Warping Method. *Journal of the Korean Society for information Management*, 36(4), 83-105. DOI : 10.3743/KOSIM.2019.36.4.083
- [15] H. Abe & S. Tsumoto. (2010). Trend detection from large text data. *2010 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 310-315. DOI : 10.1109/ICSMC.2010.5641682
- [16] M. Lee, S. Kim, H. Kim & J. Lee. (2022). Technology Opportunity Discovery using Deep Learning-based Text Mining and a Knowledge Graph. *Technological Forecasting and Social Change*, 180. DOI : 10.1016/j.techfore.2022.121718
- [17] M. N. Hwang, M. H. Cho, M. G. Hwang & D. H. Jeong. (2011). Trend Analysis of Technical Terms Using Term Life Cycle Modeling. *Korea Information Processing Society*, 18D(6), 493-500. DOI : 10.3745/KIPSTD.2011.18D.6.493
- [18] B. Fung. (2014. 04. 17.). *Hypothetical technology is fun. Real technology creeps us out*. Washington Post. <https://www.washingtonpost.com/news/the-switch/wp/2014/04/17/hypothetical-technology-is-fun-real-technology-creeps-us-out/>
- [19] S. Salvador & P. Chan. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis*, 11(5), 561-580. DOI : 10.3233/IDA-2007-11508
- [20] A. Bagnall, J. Lines, A. Bostrom, J. Large & E. Keogh. (2017). The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3), 606-660. DOI : 10.1007/s10618-016-0483-9

## 정 도 헌(Do-Heon Jeong)

[정회원]



- 2003년 6월~2017년 2월 : 한국과학기술정보연구원(KISTI) 선임연구원
- 2017년 3월~현재: 덕성여자대학교 문헌정보학과 조교수

- 관심분야 : 토픽모델링, 텍스트마닝, SNA, 유망 기술 분석
- E-Mail : doheonjeong@duksung.ac.kr

## 박 주 연(Ju-Yeon Park)

[정회원]



- 2020년 3월~현재: 덕성여자대학교 차미리사교양대학 조교수

- 관심분야 : 컴퓨팅사교력, 인공지능교육, IT융합교육
- E-Mail : juyeonpark@duksung.ac.kr