

Chapter
03다중 시계열 데이터 기반 처리 및
통합 시스템 기술

문재원_한국전자기술연구원 책임연구원

I. 결과물 개요

개발목표시기	2023. 10.	기술성숙도 (TRL)	개발 전	개발 후
			4	7
결과물 형태	SW-System, SW-Platform	검증방법	요소기술별로 자체 검증 및 공인 시험 인증 진행	
Keywords	다중 시계열 데이터 기반 처리, 다중 시계열 데이터 통합 시스템, Time-series data preprocessing, Time-series data integration system			
외부기술요소	100% 개발기술	권리성	특허, SW	

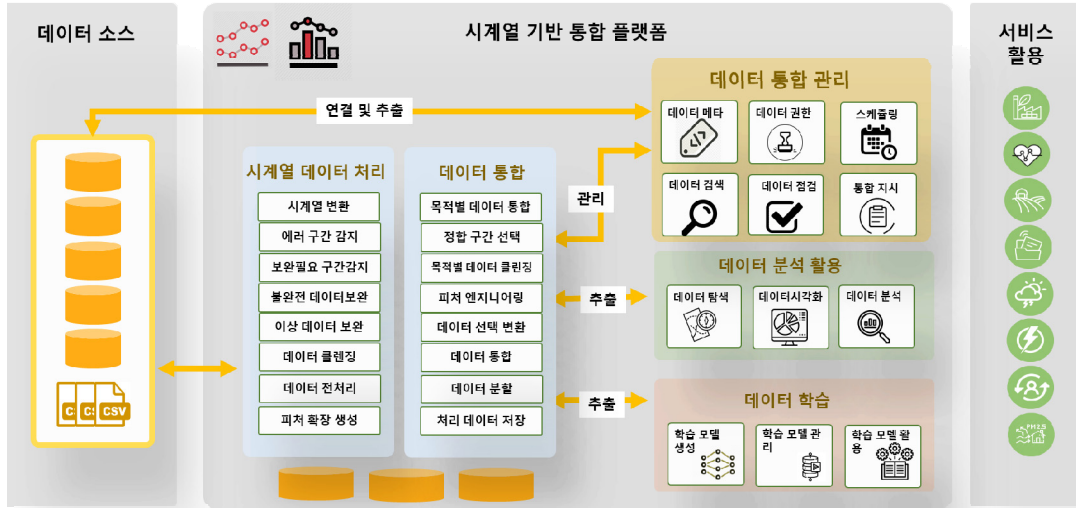
II. 기술의 개념 및 내용

- 서로 다른 구조, 양과 질, 다양한 포맷, 접근 권한, 수집 위치 등의 개별차로 파편화된 데이터를 적극 통합하고 활용하기 위한 다중 시계열 데이터 기반 처리 및 통합 시스템 기술(요소 및 플랫폼으로 구분)

* 본 내용은 문재원 책임연구원(☎ 02-6388-6657, jwmoon@keti.re.kr)에게 문의하시기 바랍니다.

** 본 내용은 필자의 주관적인 의견이며 IITP의 공식적인 입장이 아님을 밝힙니다.

***정보통신기획평가원은 현재 개발 진행 및 완료 예정인 ICT R&D 성과 결과물을 과제 종료 이전에 공개하는 "ICT R&D 사업화를 위한 기술예고"를 2014년부터 실시하고 있는 바, 본 칼럼에서는 이를 통해 공개한 결과물의 기술이전, 사업화 등 기술 활용도 제고를 위해 매주 1~2건의 관련 기술을 소개함



[그림 1] 기술개념도

III. 국내외 기술 동향 및 경쟁력

1. 기술의 특성 및 성능

- ▶ 시계열 기반 데이터 전처리·통합 패키지(SW-system)
 - 시계열 데이터의 불완전 구간 선별 기능
 - 선별된 불완전 데이터 보완 기능
 - 서로 다른 이상 패턴 감지 기능
 - 반시계열 데이터 시계열화 기능
 - 데이터 통합 및 분석 성능을 높이기 위한 기타 전처리 기능
 - 서로 다른 다형 데이터 통합 구조화 기능
 - 데이터 통합 분석 고도화를 위한 의미 확장 기능
 - 학습에 최적화된 데이터 변환 및 표상화 기능
 - 이종 도메인 시계열 데이터 분석을 위한 효과적인 지식 전달 기능
 - 기타 지엽적 다차원 다포맷 비정형 데이터 간 통합 기능

➤ 시계열 통합 데이터 분석·활용 플랫폼(SW-platform)

- 시계열 데이터 수집 및 저장 모듈
- 통합데이터 결합 활용 인터페이스
- 데이터 분석 지능 적소 배포 및 운용 기술
- 세부 서비스 사례 구축 및 사전 검증 기능
- 통합 데이터 통찰 요소기술 활용 기능
- 최종 서비스 지능 관리 및 데이터 통합 분석 기능

2. 경쟁기술/대체기술 동향 및 현황

➤ 시계열 기반 요소기술

- 빅데이터 및 AI 기술의 발전으로 분석 및 예측을 위한 요소 기술이 급속도로 발전하였으나 대부분 영상, 음성, 텍스트 데이터 포맷을 중심으로 성장하였음
- 시계열 데이터의 경우 다른 포맷에 비해 공개 데이터의 양이 적고, 품질이 보장되지 못하며 데이터마다 특성이 상이하여 요소기술 발전에 어려움이 있음
- 해외는 국내에 비해 시계열 공개 데이터와 이를 이용한 요소기술에 대한 발전이 진행되었으며, 최근 복합적이고 다양한 성질의 시계열 데이터를 변환하는 요소기술에 대한 연구가 활발히 진행되고 있음
- 국내 관련 요소 기술은 도메인을 특정한 처리 요소 기술을 중심으로 진행되거나, 스트리밍 실시간 시계열 데이터에 대한 빠른 데이터 처리 및 시각화를 위한 제품군이 출시되었음
- 해외에서는 페이스북 등의 기업에서 단변량 시계열 데이터에 대한 손쉬운 처리를 위해 소프트웨어를 공개 제공하고 있으나 다변량/복합 시계열 데이터, 반시계열 데이터에 대한 통합은 고려하고 있지 않음

➤ 시계열 기반 플랫폼 기술

- 국내 관련 플랫폼 기술은 영상, 음성, 정형화된 테이블 데이터 중심으로 상품화되었으며 시계열 기반한 플랫폼 기술 개발 역량은 IoT 기반 데이터 및 금융 데이터를 중심으로 데이터의 상황을 파악하는 시각화 모듈 위주로 개발되어 있어 전반적으로 연구 개발이

미진한 실정임

- 해외는 시계열 데이터의 빠른 처리 및 저장, 시각화와 간단한 중심으로서의 플랫폼에 대한 연구 개발이 적극적으로 진행되고 있으나 서로 다른 이질적 데이터에 대한 시계열 기반 통합 플랫폼에 대한 고려는 아직 진행된 바가 없음
 - 본 과제에서는 시계열 기반 검색/시각화 및 분석 활용 플랫폼 시장을 주요 시장으로 정의하였으며 관련 시장은 향후 급격한 규모 확장을 예상하고 있어 이에 대한 연구 개발 추진이 시급함
 - 국내에서도 AutoML을 적용한 컨셉으로 데이터를 쉽고 빠르게 융합하고 활용 학습하는 플랫폼에 대한 연구 개발이 활발하게 진행되고 있으나 시계열 데이터, 특히 서로 다른 시계열 데이터에 대한 융합 플랫폼 기술에 대한 고려는 이루어지지 않고 있음
 - 해외에서는 빅데이터에 대한 플랫폼을 구축하고 이를 클라우드 및 엣지에서 활용하려는 플랫폼 기술에 대한 연구 개발이 활발하게 진행되고 있으나, 아직은 파편적 시계열 데이터에 대한 고려는 적극적이지 않기 때문에 관련 기술에 대한 국내 기업의 기술 선점이 중요함
- 해외 공공 시계열 빅데이터
- 해외는 국내에 비해 시계열 데이터의 공개 및 연구/개발/활용 전 분야에 대해 상대적으로 더 활발하게 진행되고 있으며 특히 미국이 선도하고 있음

3. 우수성 및 차별성

경쟁기술	본 기술의 우수성/차별성
<ul style="list-style-type: none"> - 활용도메인의 다양성, 시계열 기반 알고리즘의 적극적 적용가능성, 불완전한 시계열 데이터에 대한 탐지, 이종 시계열 간 통합, 자동화, 이종 도메인 간 활용성향목 등을 전반적으로 고려하지 않고 단편적 목적에 대해서만 대응 가능함 - 국내는 해외에 비해 시계열 데이터에 대한 기술 개발이 뒤쳐진 상황이며, 해외는 Microsoft BI와 같이 시계열 데이터를 고려하는 제품군이 보다 적극적으로 상용화 및 연구 개발되고 있으나 지엽적인 이종 도메인에 대한 고려는 없는 상태임 	<ul style="list-style-type: none"> - (다양한 도메인의 요구사항 반영) 다양한 시계열 도메인에 대한 요구사항 분석 및 공통 기술 연구 개발 요소 도출 - (시계열/반시계열 데이터: 이미지, 음성, 텍스트 등의 포맷으로 절대적 시간 의존성을 갖지 않지만 시간 정보를 포함한 독립 데이터를 지칭함) - (활용 목적을 고려한 최적화된 통합) 다양한 학습 방법에 의한 모델 생성, 데이터 점검, 시각화, 구간 분석 등 서로 다른 목적을 고려한 통합 - (지엽적 데이터의 적극적 검색과 결합을 기반한 활용) 장소, 목적, 시간, 패턴 등의 유사성이나 활용 용도를 기반하여 필요한 데이터 군집을 검색하며 손쉬운 통계 응용 분야 활용 - (시계열 데이터 중심 기계 학습) 시계열 데이터에 특화된 데

경쟁기술	본 기술의 우수성/차별성
	<p>이더 분석·전처리 과정부터 실제 모델 구축까지 전 과정을 보다 쉽게 분석 및 활용 가능하게 하는 시기술에 의해 데이터 통합 분석을 수행함으로써 비전문가도 기계학습을 손쉽게 활용</p> <ul style="list-style-type: none"> - (IoT 데이터의 적극적 인공지능 활용) 시계열 데이터의 준비, 전처리, 통합, 활용, 배포에 대한 인터페이스 제공으로 전문적 활용이 용이 - (다양한 산업 도메인 요구사항 반영 및 지식 전달) 대부분 산업에서 사용되는 시계열 IoT 센서 데이터의 분석 및 활용을 일반화하여 해결 가능한 공통 정보를 추출할 수 있는 요소기술을 적용하고 플랫폼화

4. 표준화 동향

➤ 표준화 동향

- 처리 및 AI 관련 기술은 국내 TTA, 국제 ISO/IEC JTC1 등 다양한 기관에서 관련 표준화가 활발하게 진행되고 있으며, 본 과제에서 개발하는 기술에 대해서도 표준의 호환성을 고려하여 국제 표준화 기구에 적극적 참여를 통해 국제 표준을 확보해야 함

➤ 관련 보유특허

No.	국가	출원·등록번호(출원·등록일)	상태	명칭
1	대한민국	2021-0021631(2021.02.18.)	출원	파편적 시계열 데이터 검색 방법
2	대한민국	2021-0021632(2021.02.18.)	출원	파편적 데이터들의 시간축 기반 통합 방법
3	대한민국	2018-0134967(2018.11.01.) 10-2131446(2020.07.01.)	등록	분석엔진 운용 시스템, 장치 및 이의 분석엔진 갱신 방법

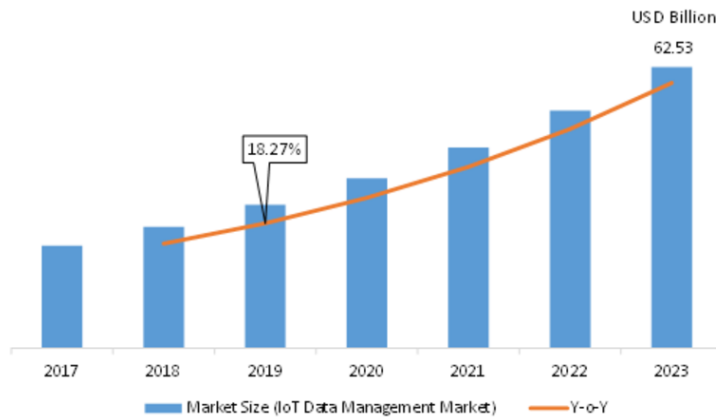
IV. 국내외 시장 동향 및 전망

1. 국내외 시장 동향 및 전망

➤ 시계열 기반 플랫폼 시장

구분	2021년	2024년	2026년
세계 IoT 데이터 처리 시장 규모	380.5억 달러	737.8억 달러	870.2억 달러
국내 IoT 플랫폼 시장 규모	9,458억 원	1조 5,437억 원	1조 7,907억 원

- 최근 Market Research Future가 발표한 자료에 따르면, 세계 IoT 데이터 처리 관련 시장규모는 2017년 228억 달러에 이르렀으며, 2023년까지 연평균 18.27% 성장하여 625억 달러에 이를 것으로 전망되고 있음



[그림 2] 세계 IoT 데이터 처리 시장규모 추이와 전망

▶ 시계열 기반 분석 시장

- 국내 빅데이터 및 분석 시장은 2023년까지 연평균성장률 11.2%를 기록하며 2조 5,692억 원의 규모에 달할 것으로 전망되며, 그 중 IT 및 비즈니스 서비스가 연평균 16.3% 성장하여 가장 큰 비중을 차지할 것임¹⁾

구분	2021년	2024년	2026년
세계 데이터 시각화 시장 규모*	107억 달러	143억 달러	174억 달러
한국 데이터 분석 솔루션 개발·공급업** 시장 규모	3,471억 원	6,763억 원	1조 550억 원

* Fortune Business Insights "Data Visualization Market Size, Share and Industry Analysis, By Component, By Solution, By Enterprise Size, By Industry and Geography Forecast, 2020-2027"

** 한국데이터산업진흥원, "2020 데이터산업현황조사"

1) 인터내셔널 데이터 코퍼레이션코리아(International Data Corporation Korea Ltd: 한국IDC) "국내 빅데이터 및 분석 2019-2023 시장 전망"

- 국내 소프트웨어정책연구소에서 발간한 2020년 소프트웨어산업 전망 보고서에 의하면, SW사업 10대 주요 이슈로 인공지능, 5G, 빅데이터 분야가 선정되었으며, eXplainable AI(설명 가능한 인공지능; xAI)는 현재 연구 초기 단계로서 테이블로 주어지는 수치 데이터 산업 영역(예; 금융)에서 적용 사례를 확대할 것으로 전망함

2. 제품화 및 활용 분야

활용 분야(제품/서비스)	제품 및 활용 분야 세부내용
스마트 산업 서비스	스마트홈, 스마트팜, 스마트 빌딩, 스마트팩토리 등 시계열 센서 데이터를 수집하는 산업군에서의 분석 예측 활용
사회 편의 서비스	금융, 부동산, 교통, 환경 문제 등 시간의 연관관계가 높은 산업군에 쌓여진 데이터들에 대한 복합 분석 및 상황 파악
복합 산업 분석 서비스	다종의 이질적 포맷의 복합적 분석이 요구되는(예; CCTV 영상과 각종 센서 등을 활용하여 안전 점검) 산업 분야에서의 활용

V. 기대효과

1. 기술도입으로 인한 경제적 효과

- 서로 다른 특성으로 인해 즉각 활용이 어려운 지엽적 데이터들이 해석되지 못한 채 공간만 차지하는 경제적 손실을 최소화하기 위해서 통합적 개별적 상황을 해석·예측하고 의사결정 활용에 연결해 데이터 처리 비용이 감소될 것으로 기대됨
- 시계열 데이터에 대한 코드의 재사용성을 높이고 배포가 용이한 형태의 데이터 파이프라인을 구축하여 개발 시간을 단축함으로써 서비스 및 애플리케이션 구축에 있어 추가 개발 리소스를 줄이는 경제적 효과를 제공하고 데이터로부터 비즈니스 가치를 극대화하도록 지원할 것으로 기대됨

2. 기술사업화로 인한 파급효과

- ▶ 여러 다양한 도메인을 고려하여 획득된 시계열 및 반 시계열 데이터를 고려하여 데이터 변환, 결합, 피쳐 생성, 데이터 정화 등의 처리 과정에 대해 재사용 가능한 코드를 생성하여 산업에서 빠르게 대응 활용할 수 있을 것으로 기대됨