

2022 KAIST Tech Fair KAIST 기술이전 설명회

선별적 밀도 갱신을 통해  
계산량이 대폭 감소한  
빅데이터 스트림 이상치 초고속 탐지 기술

전산학부 이재길 교수

KAIST



# Contents

- 01 기술/아이템 개요
- 02 연구개발 배경
- 03 기술의 특징
- 04 기술의 효과
- 05 산업분야 및 시장 규모
- 06 사업화 방안

# 1. 기술/아이템 개요 (요약 소개)

## ❖ 연구 요약

- 목적: 빅데이터 스트림에서 비정상 데이터를 (준)실시간으로 탐지하며, 특히 데이터 분포의 국소적 특성을 고려하여 보다 다양한 맥락에서의 비정상 데이터를 탐지하고자 함
- 주요 아이디어: 커널 밀도 추정 (KDE) 기반 선별적 상대 밀도 갱신
  - 데이터 스트림을 슬라이딩 윈도우로 관리하며 KDE 기반 상대적 지역 밀도 추정 및 이상치 점수화
  - 변화가 많이 발생하는 국소 지역만 빠르게 선별하여 밀도 분포 및 이상치 점수 갱신을 최소화
- 기대 효과: 스마트 팩토리의 제조 공정, 웨어러블 헬스케어 기기, 클라우드 서비스, 모빌리티 서비스 등과 같이 다양한 실제 응용 분야에서 발생하는 빅데이터 스트림에서 빠르고 정확하게 이상치(불량품, 위험 상황 등) 탐지 및 잠재적 이상치 모니터링

## ❖ 주요 성과

- 네트워크 트래픽, 모션 센서 데이터, 심전도 등 6종의 벤치마크 데이터에서 우수 성능 입증
  - 기존 방법 대비 11~3,107배 빠른 속도로 지역 이상치 탐지
  - 기존 방법과 비슷한 수준의 메모리 사용량, 높거나 비슷한 수준의 탐지 정확도 달성

## 2. 연구개발 배경

### ❖ 연구 배경 및 필요성

- 빅데이터 스트림에서 비정상 데이터를 신속 정확하게 탐지하는 일은 여러 실시간 응용에서 매우 중요

#### 데이터 소스



스마트 팩토리

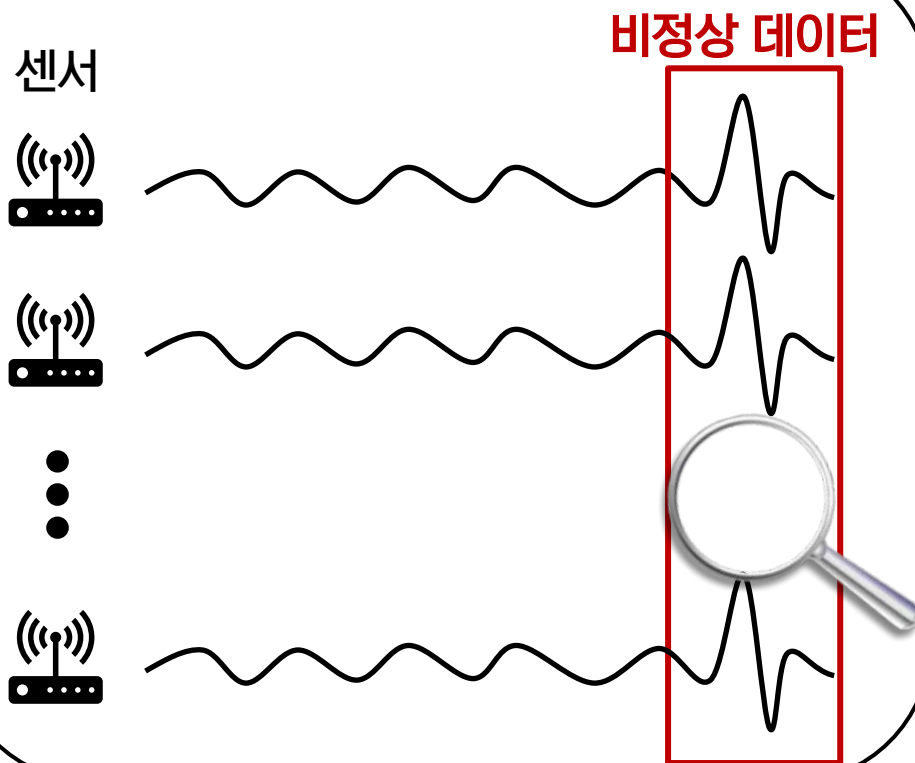


헬스케어

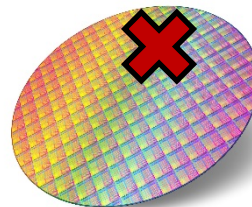


모빌리티

#### 데이터 스트림



#### 탐지 이벤트



불량 제품



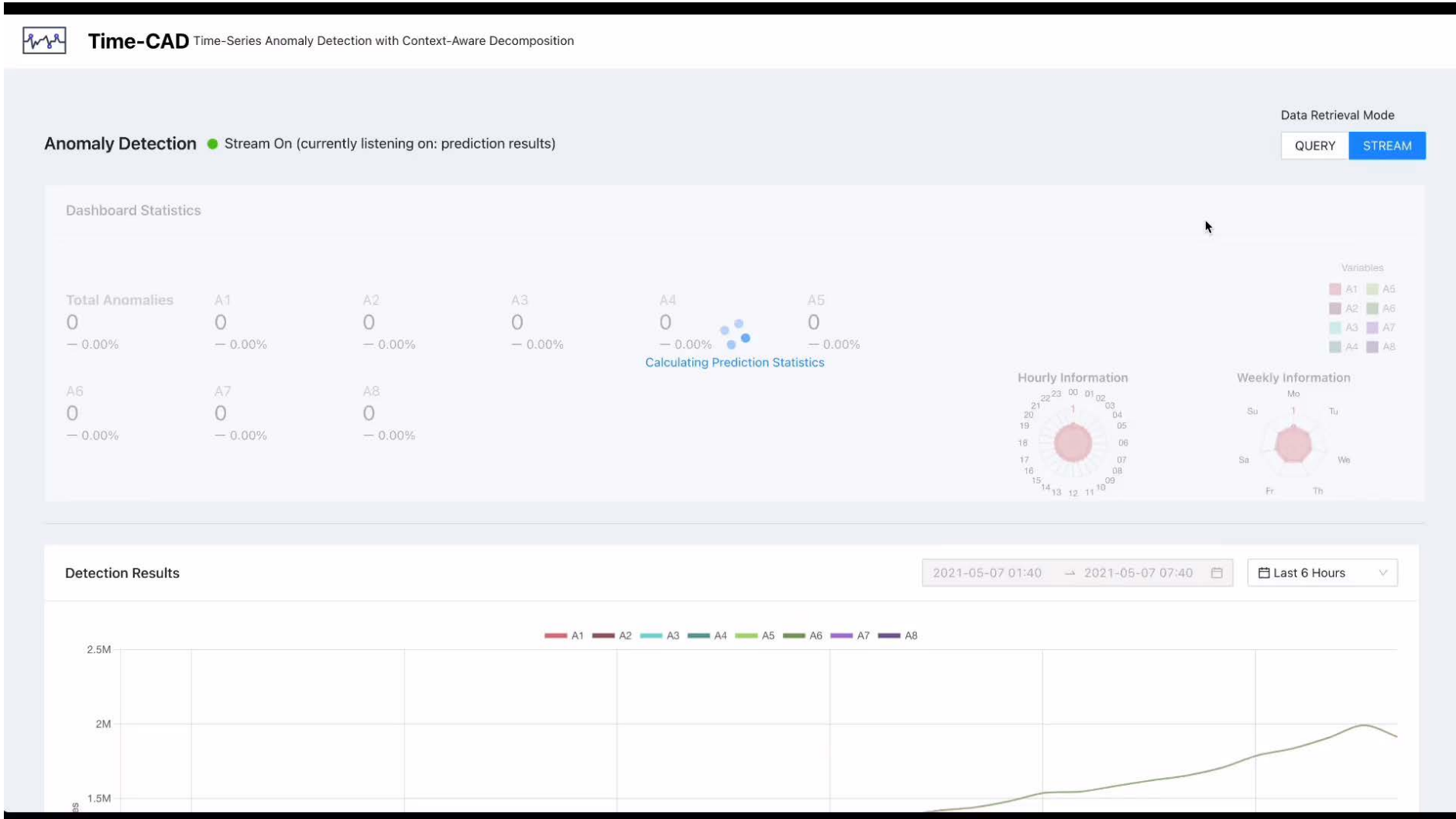
응급 상황



사고 위험

## 2. 연구개발 배경

### ❖ 예시

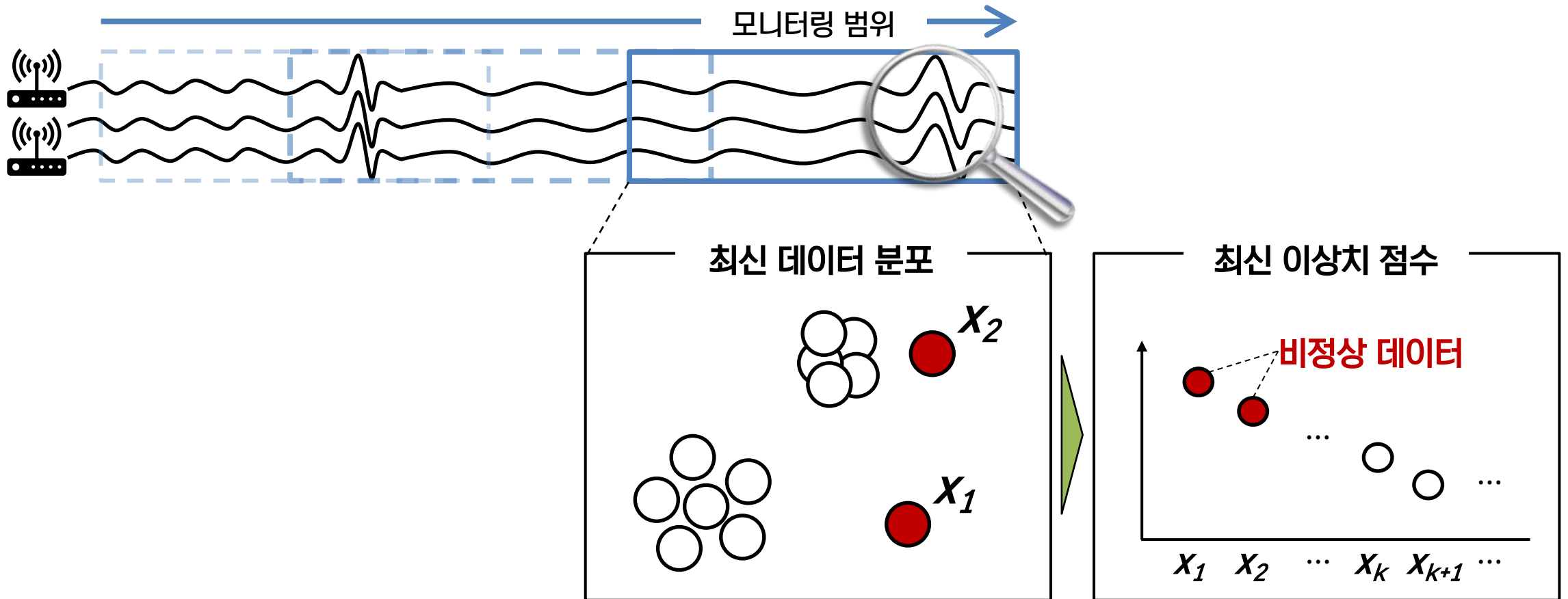




### 3. 기술의 특징 (1/3)

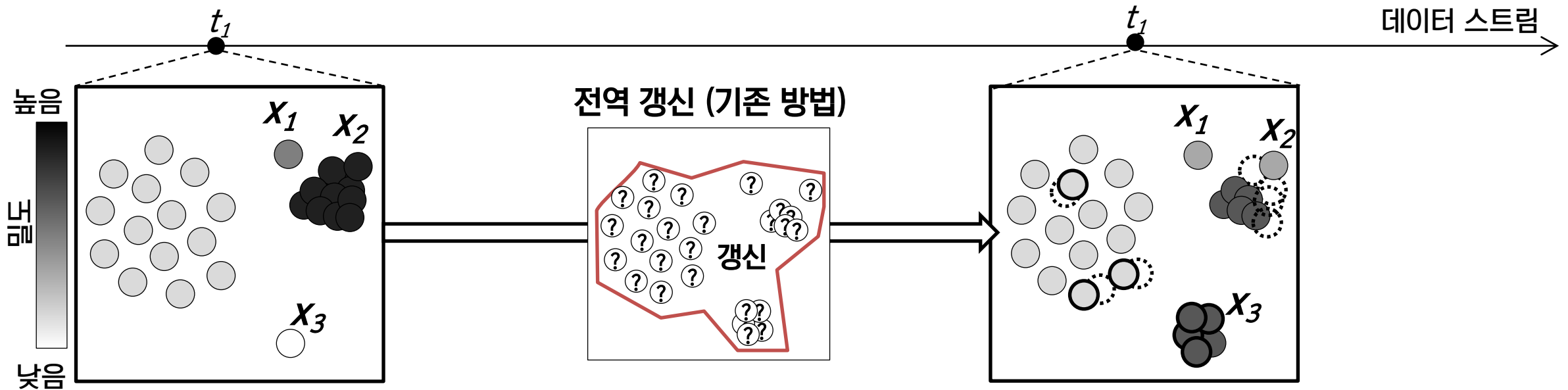
#### ❖ 연구 요약

- 빅데이터 스트림에서 **불변 지역 스키핑** 기법을 통해 **신속 정확하게 이상치 점수 갱신**



## ❖ 기존 방법의 한계

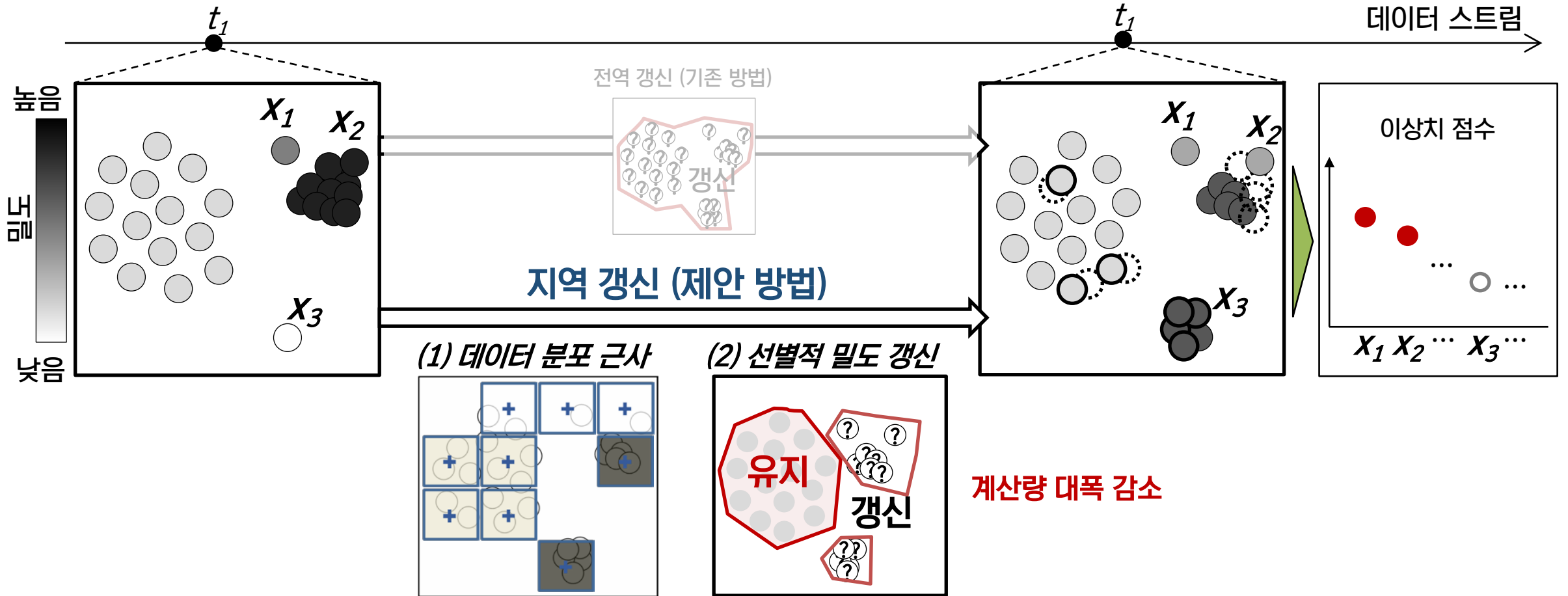
- 기존 방법은 실제 변화 정도에 관계없이 매번 **데이터 밀도 분포**를 **전역 갱신**하여 **비효율적**



### 3. 기술의 특징 (3/3)

#### ❖ 핵심 기술

- 제안 방법은 최적의 **데이터 분포 근사**를 통해 변화가 많이 발생하는 데이터에 한해 **선별적 갱신**

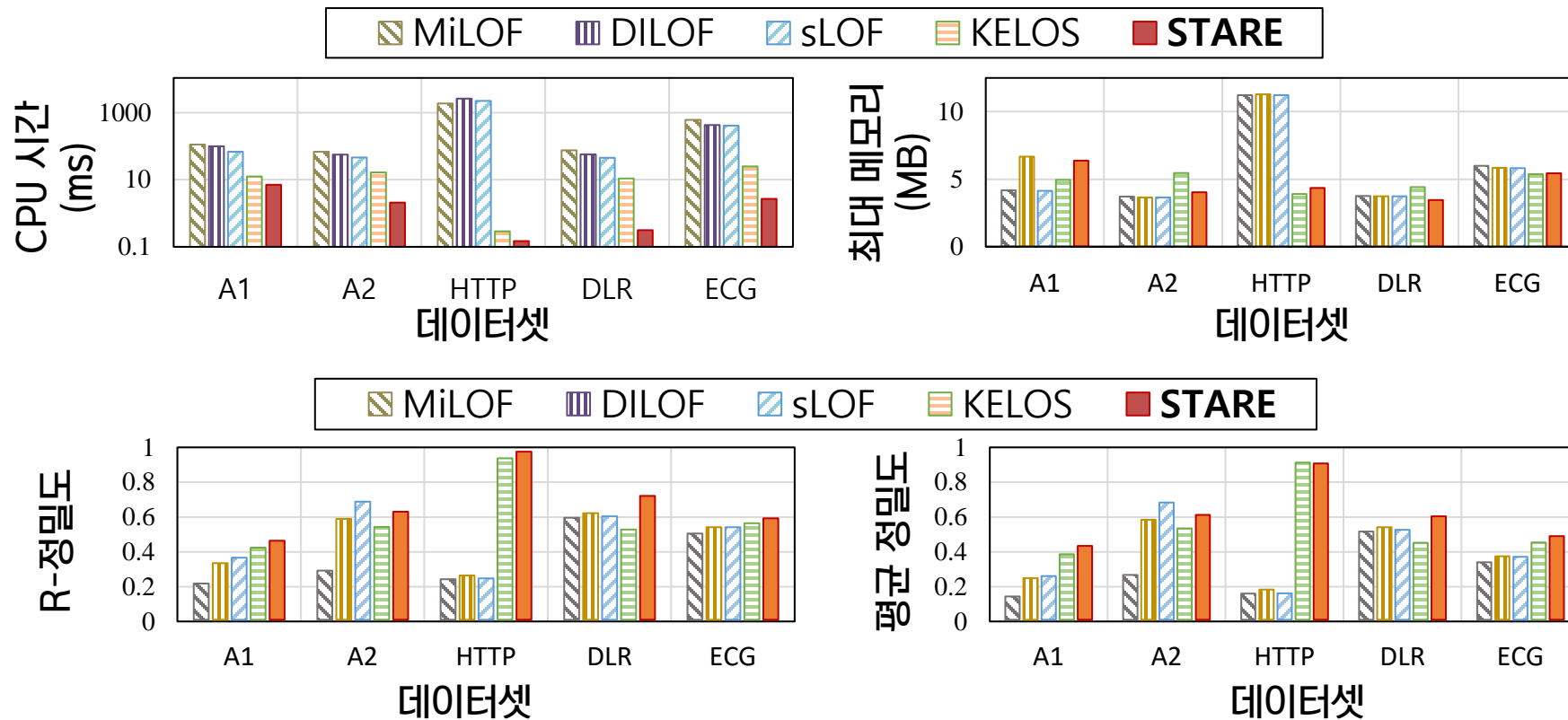




## 4. 기술의 효과

### ❖ 네트워크 트래픽(A1/2, HTTP), 모션 센서(DLR), 심전도(ECG) 등 6종의 벤치마크 데이터 성능 비교

- 제안 알고리즘(STARE)이 기존 알고리즘보다 **11~3,107배 빠른 속도로 정확하게 이상치 탐지**



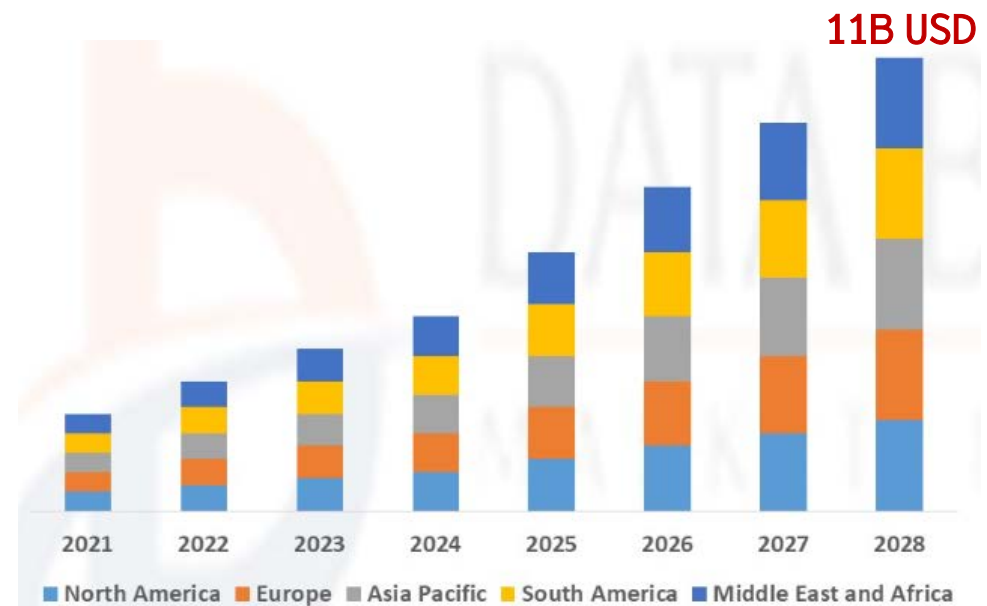
## 5. 산업분야 및 시장 규모

### ❖ 산업 분야

- 실시간 데이터가 수집되고 이를 모니터링해야 하는 모든 산업군에서 필요
- 예: Microsoft 등의 Cloud 서비스, 삼성전자 등의 제조업, 안랩 등의 사이버 보안 ...

### ❖ 시장 규모

- 전세계 이상치 검출(anomaly detection) 시장 규모는 2028년에 약 110억 달러까지 증가할 전망



출처: <https://www.databridgemarketresearch.com/reports/global-anomaly-detection-market>

- ❖ 대한민국 특허: 다차원 데이터 스트림에서의 실시간 이상치 탐지 방법 및 장치, 제10-2246303, 2021년 4월 23일
- ❖ 특허 기술이전(통상 실시권) 희망
- ❖ 도입기업 자체 SW에 본 특허 기술 통합 가능
- ❖ 소스코드 제공 및 교수 자문 가능