

이상탐지 A to Z

7편.

비지도 학습 기반의 머신러닝 기법을 활용한 이상탐지(2부)

데이크루 2기 Team Zoo

목차

1. ICA
2. ICA 와 CLT
3. ICA와 PCA의 차이
4. ICA 이상탐지 사례



1.ICA

ICA모델

독립 성분 분석(Independent Component Analysis, ICA)

다변량의 신호를 통계적으로 독립적인 하부 성분으로 분리하는 계산 방법

ICA example

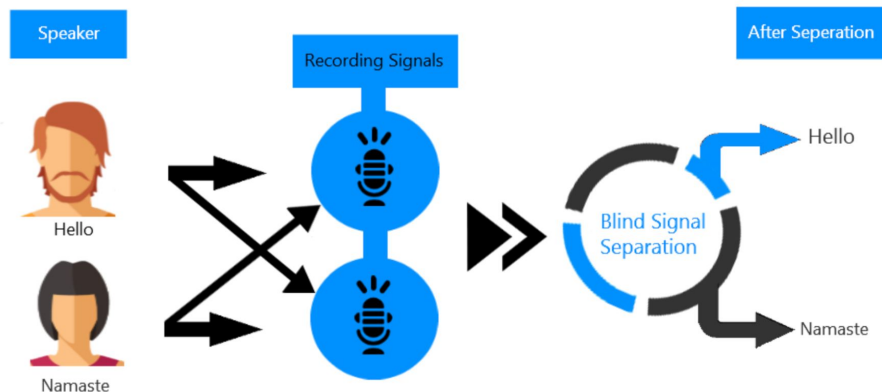
섞여버린 두 개의 음원을 분리해내는 과정
Blind Source Separation = ICA 분석

-> 다변량 신호 가지고 독립적인 성질로
분리 역할

$$x_1(t) = a_{11}s_1(t) + a_{12}s_2(t)$$

$$x_2(t) = a_{21}s_1(t) + a_{22}s_2(t)$$

$$s = A^{-1}x = Wx$$



목적: 두 녹음 음원으로부터 음원을 분리 ($s_1(t)$, $s_2(t)$ 각각의 변수로 분리)

행렬을 이용하여 W 행렬 찾아 source 구해야한다.



2.ICA와 CLT

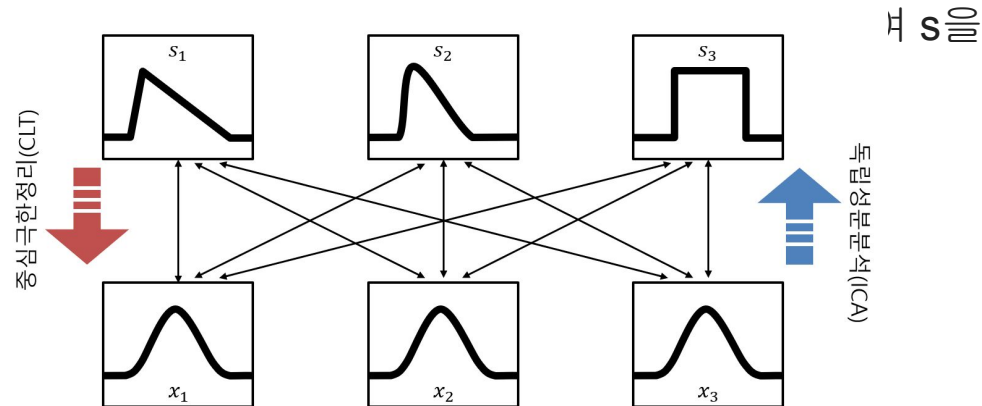
중심극한정리(CLT) 개념 <-> 독립성분분석(ICA) 개념

1)CLT: x 을 찾는 것 목적

- 서로 독립적인 랜덤 변수들의 선형조합으로 이루어진 새로운 랜덤 변수 분포는 가우스 분포를 따른다.
- 선형 조합에 들어가는 독립변수 많음 -> 가우스 분포에 가까워짐

2)ICA (독립성분분석): 중심극한정리(CLT)와 독립성분분석(ICA) 관계 개요도

- 독립 랜덤 변수들 찾는 과정이다.



→ 서로 독립적인 랜덤 변수들의 분포의 선형조합은 가우스 분포를 따른다.

→ (s 에 비해 더 가우스 분포를 따르는) x 들을 어떻게 조합하면 s 를 얻을까?



3.ICA와 PCA 차이

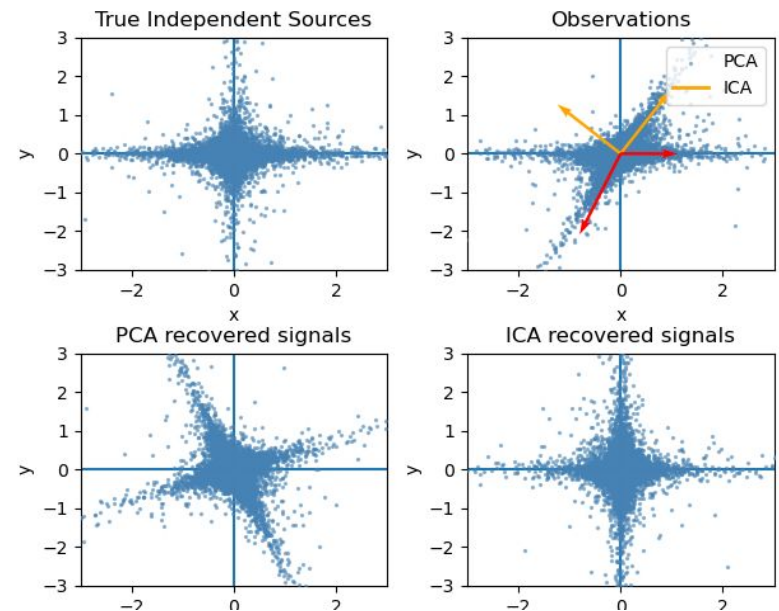
-공통점: 주어진 데이터를 대표하는 기저벡터를 찾아준다.
(같은 역할)

1)PCA:

- feature space에서 **직교**하는 기저 벡터 집합
- 기저벡터 우선순위: 정사영시 **분산이 최대인** 벡터

2)ICA:

- 기저벡터들이 서로 **직교하지 않을** 수도 있다.
- 기저벡터 우선순위: 그 결과들이 최대한 **독립적**



4.ICA 이상탐지 사례

ICA을 이용한 다변량 공정에서 공장탐지 방법

MSPM 방법 사용 :

- 다변량 공정에서 고장탐지 위해 통계적 온라인 모니터링 사용
- 비정상 행동 사전에 탐지
- PCA와 ICA기법이 대표적인 고장탐지 모델

고장 탐지 절차

-1.z 점수 표준화:

범위 동일하게 맞춤

-2.화이트닝, 화이트닝 변환

독립성분들을 추출하고 독립성분 개수 설정

-3.통계량 문턱값 설정

: 비정상인지 정상인지 판별하기 위한 문턱값

-4. SPE 통계량 계산후 시스템 고장 여부 판별하기(기준: 문턱값)

Table. 1 Comparison of fault detection performance.

Data	ICA				PCA			
	I_d^2		SPE		T^2		SPE	
	T1	T2	T1	T2	T1	T2	T1	T2
Bias data	2.1	8.2	2.2	88.9	2.14	71.1	1.7	13.1
Drift data	5.8	38.5	2.6	89.9	3.1	76.8	3.5	43.2
Feedwater heater	4.4	16.3	0	68.1	4.3	38.3	1.5	18.7

(I^2 :차원 축소된 독립성분 통계량,SPE: 근사오차 크기)



참고 자료

<https://gowrishankar.info/blog/cocktail-party-problem-eigen-theory-and-blind-source-separation-using-ica/>
<https://angeloyeo.github.io/2020/07/14/ICA.html>
<https://www.koreascience.or.kr/article/JAKO202005653789506.pdf>

