

Data Mining and Machine Learning LAB

Data Mining and Machine Learning LAB



✉ 박정희 교수님 (연구실: 520호)

cheonghee@cnu.ac.kr

✉ 대학원생 연구실: 528호

✉ 연구분야

Machine Learning, Deep Learning, Data Mining

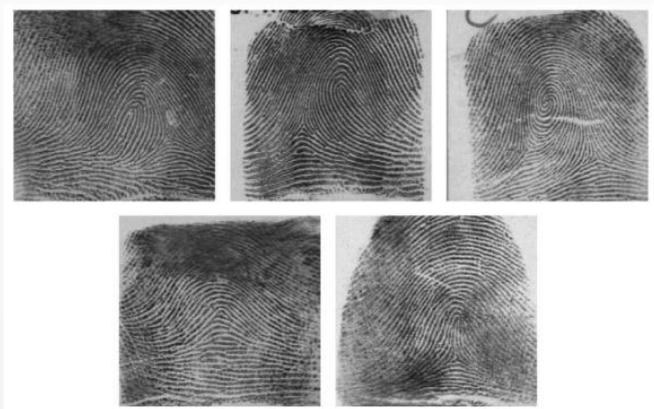
✉ 랩 홈페이지: <http://dmml.cnu.ac.kr>

연구 주제

- 기계학습, 딥러닝, 데이터 마이닝 등 연구분야에서 다양한 주제를 연구
- 분류, 군집화, 이상 탐지, 차원 감소 등의 기계학습, 딥러닝, 데이터마이닝 알고리즘 개발
- 응용분야에서 발생하는 문제에 기계학습, 딥러닝, 데이터마이닝 알고리즘을 적용하여 문제 해결
- 2000년대 초반부터 현재까지 수행했던 연구주제들 중에서 몇가지를 살펴보도록 하겠습니다.

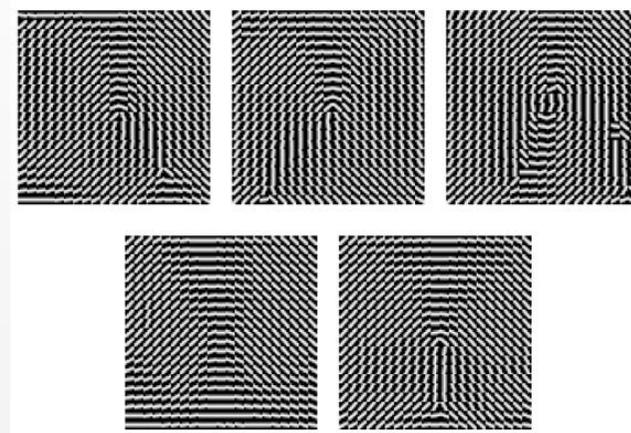
랩에서 수행한 연구 주제 – 2005년

- 지문 분류 (Fingerprint classification) 알고리즘 개발



5가지 종류의 지문 이미지:

Discrete Fourier transform과
Directional filter 적용



Directional image

차원 감소

분류기에 의한 5가지 유형 지문 분류 수행

- 논문: Fingerprint Classification using Fast Fourier Transform and Nonlinear Discriminant Analysis

랩에서 수행한 연구 주제 – 2008년

- Multi-labeled problem에 대한 분류 알고리즘 개발
 - Multi-labeled problem이란?

While traditional classification problem assumes that a data sample belongs to only one class among the predefined classes, a multi-labeled problem can arise in real situation where a data sample is associated with multiple class labels. For example, in text categorization documents can be classified to multiple categories of topics (Lewis et al., 2004). In bioinformatics, each gene is associated with a set of functional classes (Pavlidis et al., 2001). In a multi-labeled problem, the main task is to output a set of class labels associated with a new unseen data sample. One com-

- 논문: On applying linear discriminant analysis for multi-labeled problems 등

랩에서 수행한 연구 주제 – 2008년

- 다양한 조명 조건하에서 얼굴인식 문제



Figure 1: Face images under various illumination conditions.



Figure 4: comparison of face images reconstructed by the proposed methods

- 논문: An efficient image normalization method for face recognition under varying illuminations

랩에서 수행한 연구 주제 – 2009년

- SVM을 이용한 데이터 전처리(discretization)와 associative classification

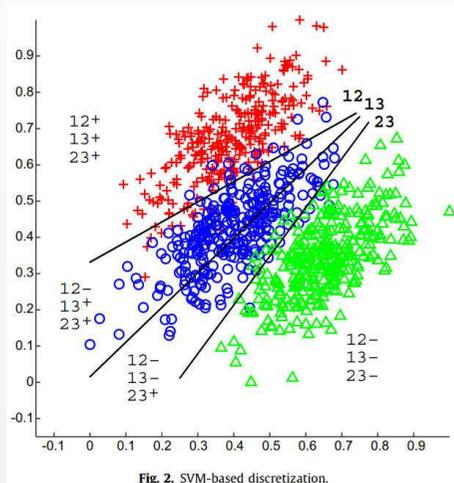
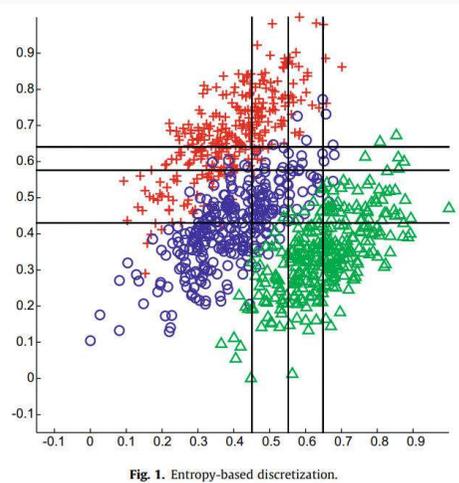


Table 2

Comparison of error rates (%) by different discretization methods

	Entropy-based disc.	SVM-based disc.	SVM+entropy-based disc.
CBA	5.3	2.2	2.1
C4.5	4.6	2.3	1.7

Some rules generated by CBA using entropy-based discretization

- rule1: health = not_recom → class = not_recom
- rule2: $0.55 < x < 0.65$ and $y < 0.43$ → class = spec_prior
- rule4: $x > 0.65$ and health = priority → class = spec_prior

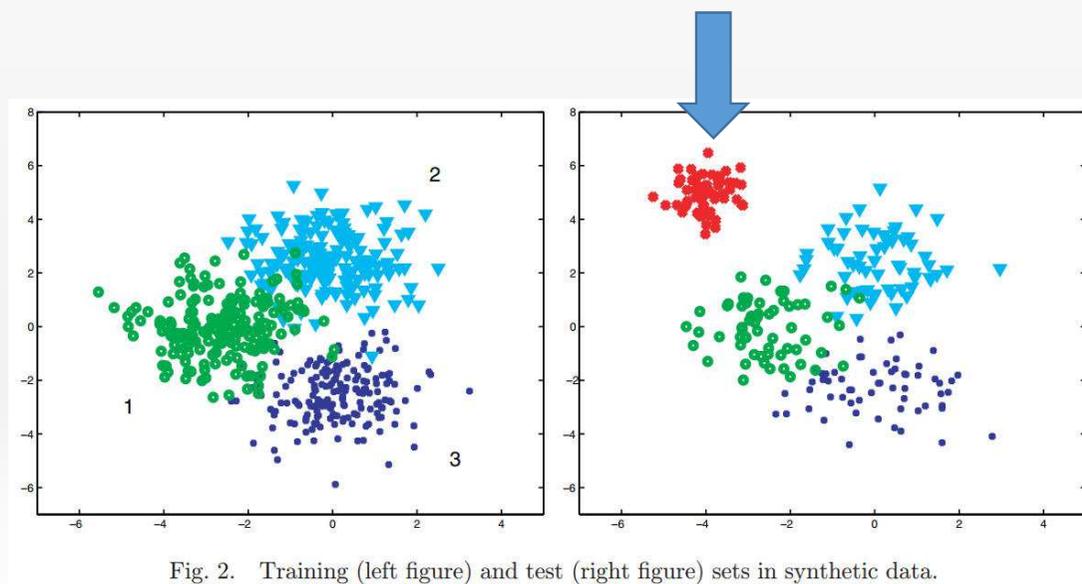
Some rules generated by CBA using SVM-based discretization

- rule1: health = not_recom → class = not_recom
- rule2: $0.6x - y + 0.3 > 0$ and health = recommended → class = priority
- rule3: $1.3x - y - 0.3 < 0$ and has_nurs = critical → class = spec_prior

- 논문: A SVM-based discretization method with application to associative classification

랩에서 수행한 연구 주제 – 2010년

- 학습단계에서 존재하지 않았던 새로 나타난 클래스를 탐지할 수 있을까?
- Hypothesis testing for change detection in class distribution
 - > Alarm for Emerging new class
 - > Retrieval of new class members



- 논문: Detection of emerging new class using statistical hypothesis testing and density estimation

랩에서 수행한 연구 주제 – 2015년

- 허밍(humming: 노래 흥얼거림)에 의한 음악 검색 방법
- Spectral Hashing과 동적 타임 워핑(dynamic time warping) 방법 적용

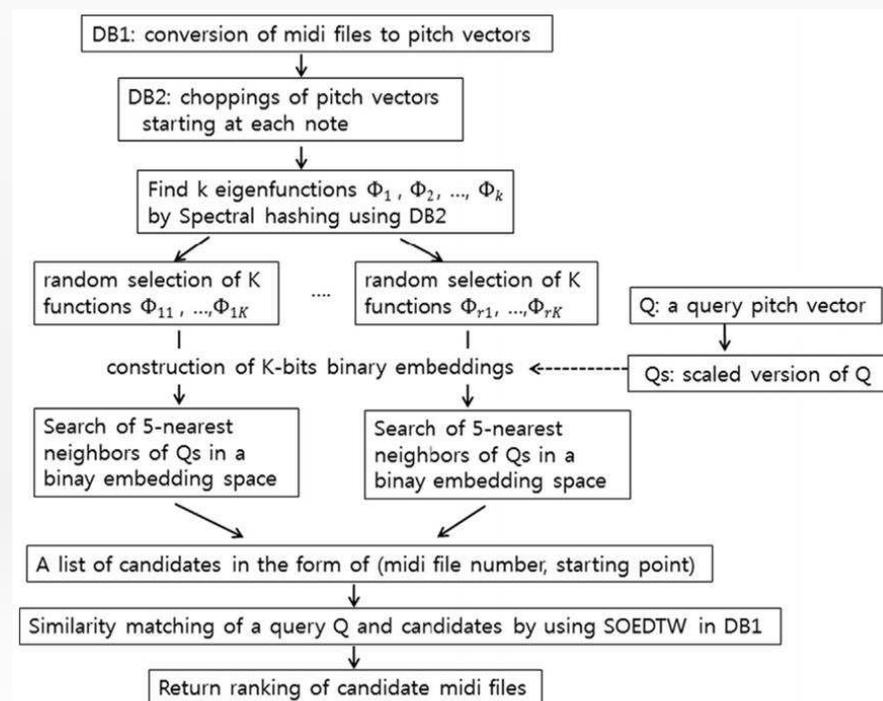
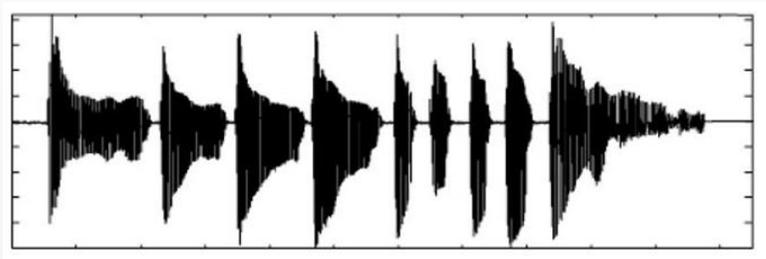


Fig. 1. A flowchart of the proposed QBH method.

- 논문: Query by humming based on multiple spectral hashing and scaled open-end dynamic time warping

랩에서 수행한 연구 주제 – 2017년

- 트위터 스팸 탐지

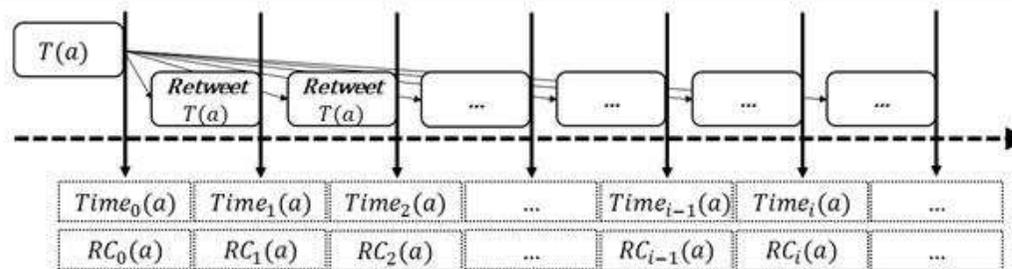


Figure 2. The occurrence of retweets from an original tweet

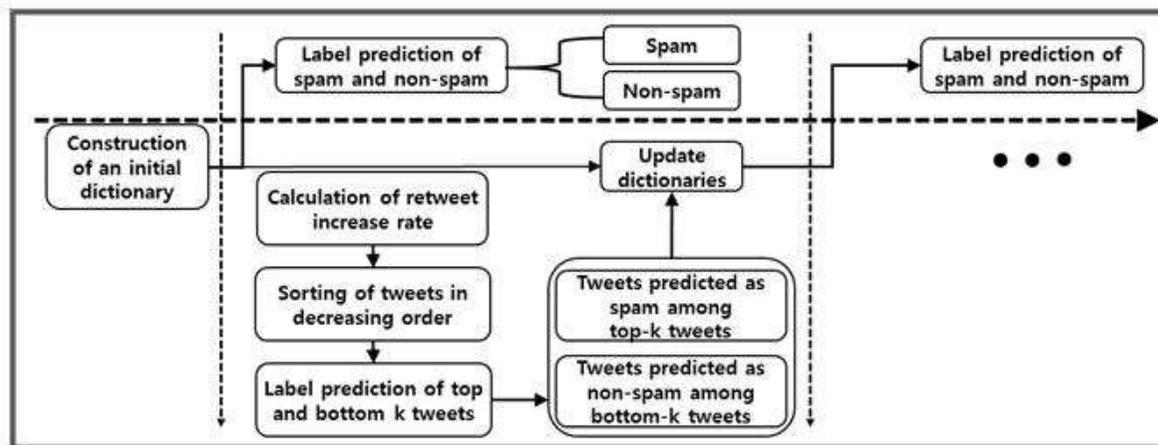
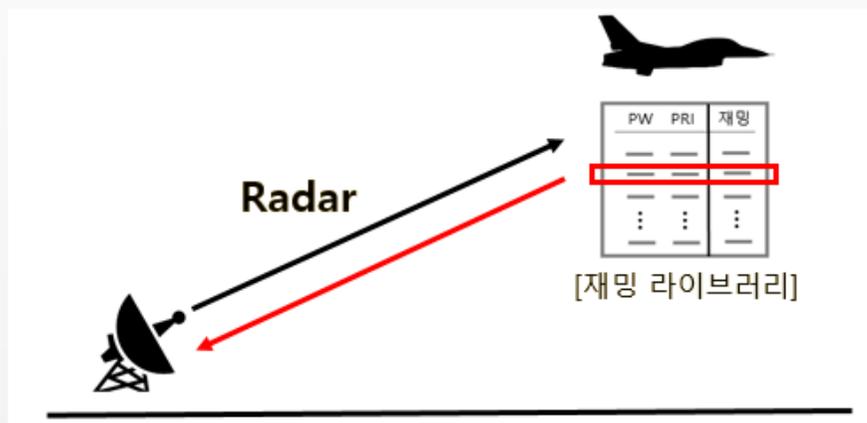


Figure 3. A flowchart of the proposed method when an initial training set is small

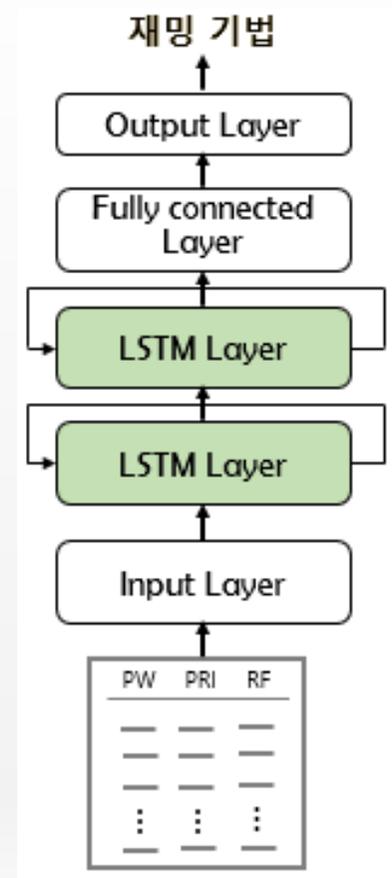
- 논문: 트레이닝 데이터가 제한된 환경에서 N-gram 사전을 이용한 트위터 스팸 탐지 방법

랩에서 수행한 연구 주제 – 2019년

- LSTM을 이용한 레이더 재밍 기법 예측



라이브러리를 이용하는 기존의 재밍 방법



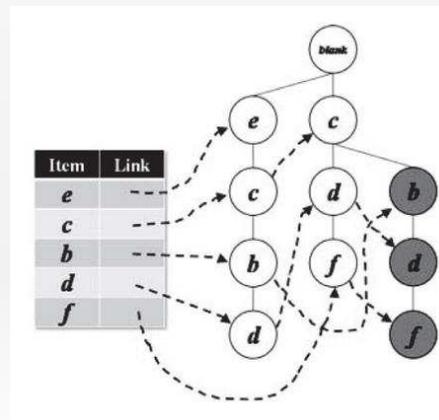
LSTM을 이용한 재밍 방법

- 논문: Jamming Prediction for Radar Signals Using Machine Learning Methods

랩에서 수행한 연구 주제 – 2019년

- 트위터 스트림에서 새롭게 부상하는 핫 토픽 탐지
 - High Utility Pattern Mining과 Topic tree를 이용하여 핫 토픽을 나타내는 단어 군집 탐지

Twitter Stream $\underbrace{T_1 T_2 T_3}_{B_1} \underbrace{T_4 T_5 T_6 T_7}_{B_2} \dots \underbrace{T_i T_{i+1} T_{i+2}}_{B_L} \dots$



- 논문: Emerging topic detection in twitter stream based on high utility pattern mining

랩에서 수행한 연구 주제 – 2020, 2021년

- 스트림 데이터에서 실시간 이상 패턴 탐지 알고리즘 개발과 적용
 - 고전압 지중 전력 케이블에서 부분 방전 탐지
 - 스마트 계량기 데이터에서 전기 계약종별 위반 탐지



- 논문: - Anomaly pattern detection for streaming data
 - 전력 소비 데이터 스트림에서 이상 패턴 탐지 방법을 이용한 전기 계약종별 위반 탐지
 - Partial Discharge Detection Based on Anomaly Pattern Detection

연구실 졸업생들은 어디에?

랩에서 수행한 연구 주제 – 2010년

- 학습단계에서 존재하지 않았던 새로 나타난 클래스를 탐지할 수 있을까?
- Hypothesis testing for change detection in class distribution
 - > Alarm for Emerging new class
 - > Retrieval of new class members

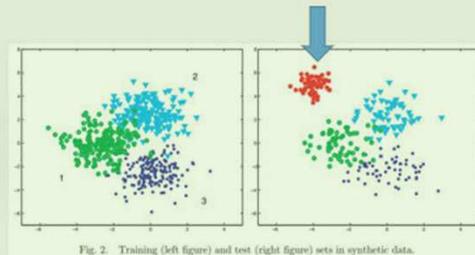


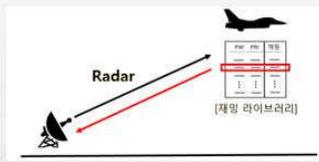
Fig. 2. Training (left figure) and test (right figure) sets in synthetic data.

• 논문: Detection of emerging new class using statistical hypothesis testing and density estimation

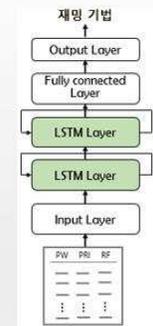
이 연구를 했고
현재 삼성탈레스 재직

랩에서 수행한 연구 주제 – 2019년

- LSTM을 이용한 레이더 재밍 기법 예측



라이브러리를 이용하는 기존의 재밍 방법



LSTM을 이용한 재밍 방법

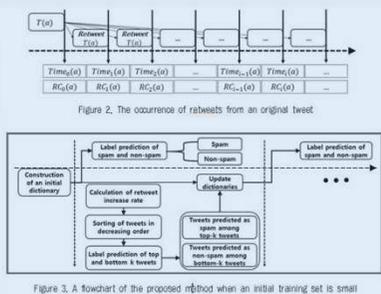
• 논문: Jamming Prediction for Radar Signals Using Machine Learning Methods

이 연구를 했고
현재 엔씨소프트 재직

연구실 졸업생들은 어디에?

랩에서 수행한 연구 주제 - 2017년

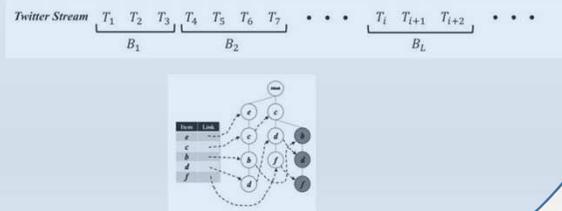
- 트위터 스팸 탐지



- 논문: 트레이닝 데이터가 제한된 환경에서 N-gram 사전을 이용한 트위터 스팸 탐지 방법

랩에서 수행한 연구 주제 - 2019년

- 트위터 스트림에서 새롭게 부상하는 핫 토픽 탐지
 - High Utility Pattern Mining과 Topic tree를 이용하여 핫 토픽을 나타내는 단어 군집 탐지



- 논문: Emerging topic detection in twitter stream based on high utility pattern mining

이 연구를 했고 석사 졸업 후 현재 삼성전자 재직

졸업생 (2021년까지)

박사 3명: 전자통신연구원, ADD

석사 18명: NC소프트, 삼성전자, ADD(국방과학연구소)

한국전력연구소, 삼성탈레스,

국가보안기술연구소, ETRI(전자통신 연구원) 등

연구 장학생 문의 환영합니다.

- 기계학습, 딥러닝, 데이터 마이닝 등 연구분야에서 다양한 주제를 연구
- 분류, 군집화, 이상 탐지, 차원 감소 등의 기계학습, 딥러닝, 데이터마이닝 알고리즘 개발
- 응용분야에서 발생하는 문제에 기계학습, 딥러닝, 데이터마이닝 알고리즘을 적용하여 문제 해결
- 같이 연구를 진행할 학부 연구원과 대학원 입학 관련 문의 환영합니다!!
- 이메일: cheonghee@cnu.ac.kr