

Natural Language Processing in deep learning

A&O Lab 최정윤

목차

1. Machine Learning
2. Deep Learning
3. Natural language processing in deep learning
 - a. Motivation
 - b. Natural Language Processing
 - c. Word Representation
 - d. Sentence Representation
 - e. Text Classification
 - f. Language Model
 - g. Text to speech

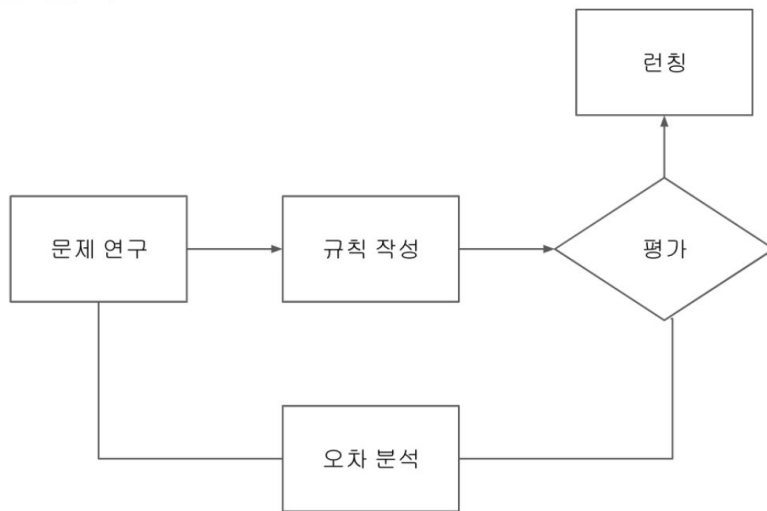
1. Machine Learning

Reference 1. machine learning

Reference: 기계학습 개론(2015, 서울대학교 강의)

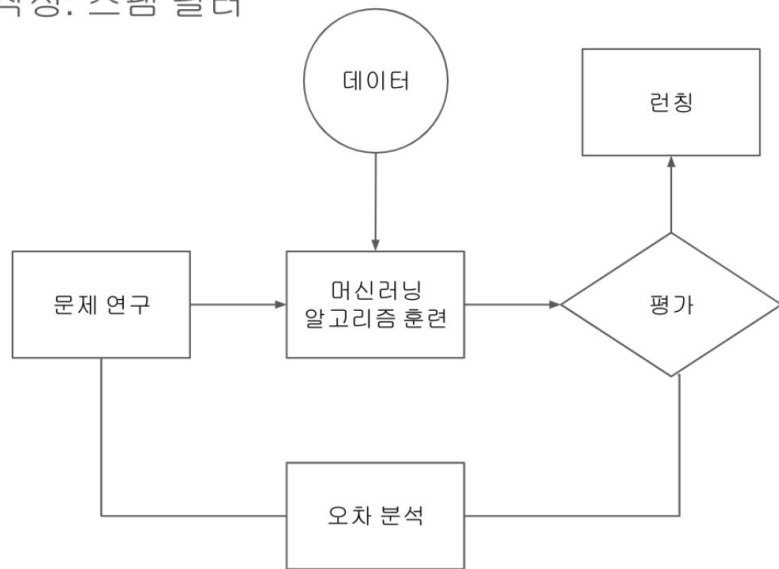
1. Machine Learning

기존 프로그램 작성: 스팸 필터



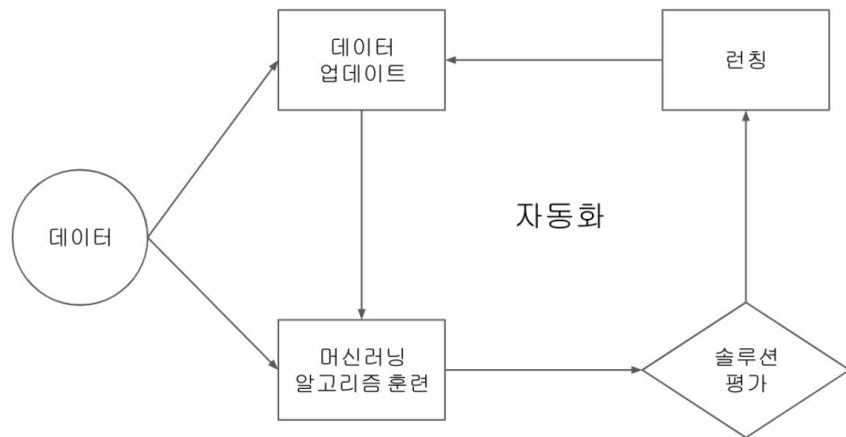
1. Machine Learning

머신러닝 프로그램 작성: 스팸 필터



1. Machine Learning

머신러닝 프로그램 작성: 스팸 필터

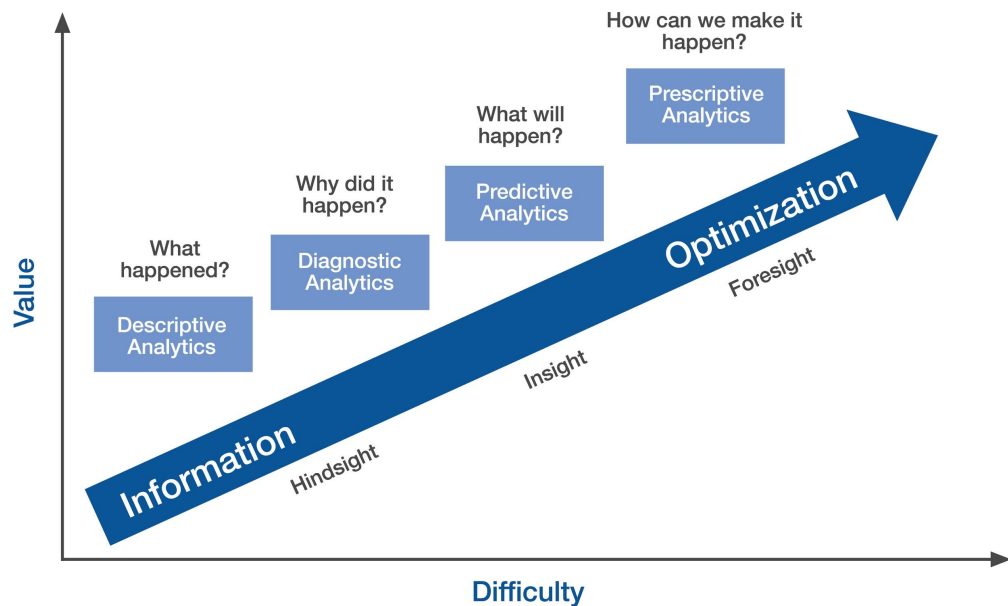


1. Machine Learning

- 1) Information based learning
 - decision tree
- 2) Similarity based learning
 - KNN
- 3) Probability-based learning
 - Bayesian Network
- 4) Error-based learning
 - Regression, SVM

1. Machine Learning

Predictive Analytics



작은 예측이 커다란 효과를 가져온다.

“기업들은 이미 매스 마케팅이라는 일종의 숫자 게임을 하고 있다. 그렇기에 예측을 통해서 게임의 균형을 미묘하게 그러나 의미 있는 수준 안에서 자기에게 유리한 쪽으로 기울어지게 할 수 있으며, 설사 예측이 매우 정확하지 않아도 그렇게 할 수 있다. **사실 예측의 효용성은 꽤 낮은 정확도에도 불구하고 유의미하다.** 어렵פות하게라도 볼 수 있는 것이 완전히 깜깜한 암흑 상태보다는 우수한 결과를 낳는다.”

Reference: 빅데이터의 다음 단계는 예측 분석이다. (p38)

2. Deep Learning

Reference 2. Deep Learning

2. Deep Learning

딥러닝에서 모든 것은 벡터이다.

모두가 기하학적 공간에 놓인 하나의 포인트이다.

데이터가 가진 의미를 벡터와 기하학적 공간으로 변환한 후 한 공간에서 다른공간으로 매핑하는 복잡한 기하학적 변환을 점진적으로 학습한다.

2. Deep Learning

딥러닝의 한계

- 딥러닝 모델은 한 벡터 공간을 다른 벡터 공간으로 매핑하기 위해 단순하고 연속된 기하학적 변환을 연결한 것이다.
- 현재 딥러닝 기술로 달성하기 어려운 애플리케이션이 많다.
- 제품 관리자가 작성한 소프트웨어 기능 정의서와 개발 팀이 이 요구 사항에 맞추어 개발한 소스 코드로 이루어진 데이터를 수십 만개 모을 수 있어도 제품 설명서를 보고 소스 코드를 생성하는 딥러닝 모델은 훈련할 수 없다.

3. Natural Language Processing

3-1. Motivation

서시, 윤동주

죽는 날까지 하늘을 우르러
한점 부끄럼이 없기를,
앞새에 이는 바람에도
나는 괴로워했다.
별을 노래하는 마음으로
모든 죽어가는것을 사랑해야지
그리고 나한테 주어진 길을
거러가야겠다.

오늘밤에도 별이 바람에 스치운다。

3. Natural Language Processing

3-1. Motivation

별 헤는 밤, 윤동주

별 하나에 추억과,
별 하나에 사랑과,
별 하나에 쓸쓸함과,
별 하나에 동경과,
별 하나에 시와,
별 하나에 어머니, 어머니,

3. Natural Language Processing

3-2. Natural Language Processing

- Natural Language Processing is hard.
- Tokenization.
- Classical Natural Language Processing.

3. Natural Language Processing

3-3. Word Representation

- Sparse
- Dense
 -
 -
 -

[Word2Vec](#)

Representation

Representation

[Example](#)

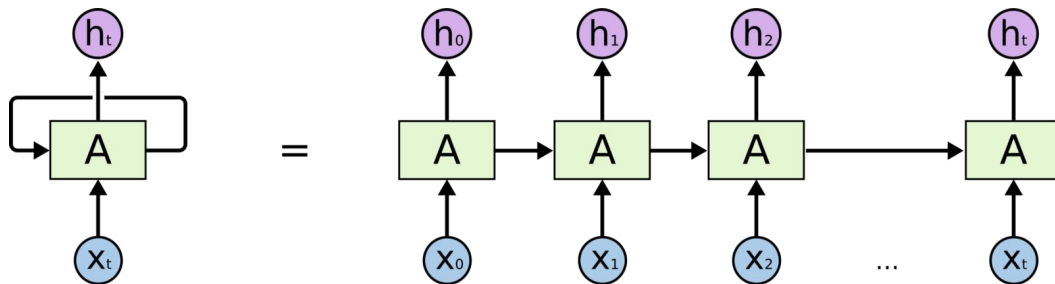
Glove

[FastText](#)

3. Natural Language Processing

3-4. Sentence Representation

- Bag Of Words - Count, TF/IDF
- Continuous Bag of Words
- Convolutional Network
- Recurrent Network
- Self-Attention Network



3. Natural Language Processing

3-5. Text Classification

- Reference 참고

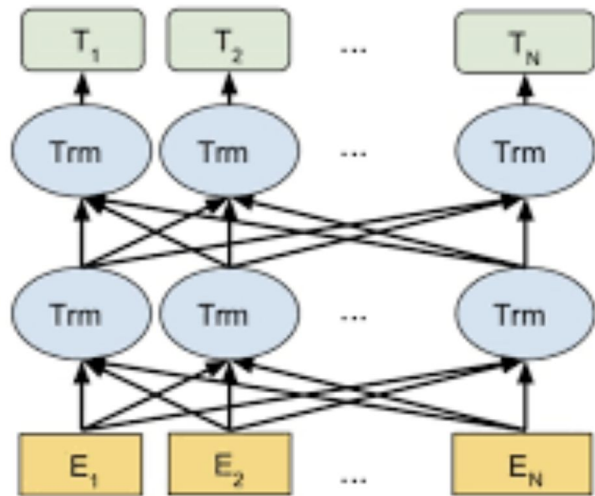
- <https://fasttext.cc/docs/en/supervised-tutorial.html>

3. Natural Language Processing

3-6. Language Model

- 언어를 생성하는 모델
 -
 -
 - Neural Language
- Neural Machine Translation
- Question & Answering

BERT (Ours)



ile
tic

Reference:

- 1) <https://www.edwith.org/deepnlp/joinLectures/17363> 2) <https://github.com/oxford-cs-deepnlp-2017/lectures>

3. Natural Language Processing

3-7. Text to speech

- text to speech



Reference: 1) <https://github.com/oxford-cs-deepnlp-2017/lectures>