

NLP Introduction

강사 : 백병인

pi.paek@modulabs.co.kr

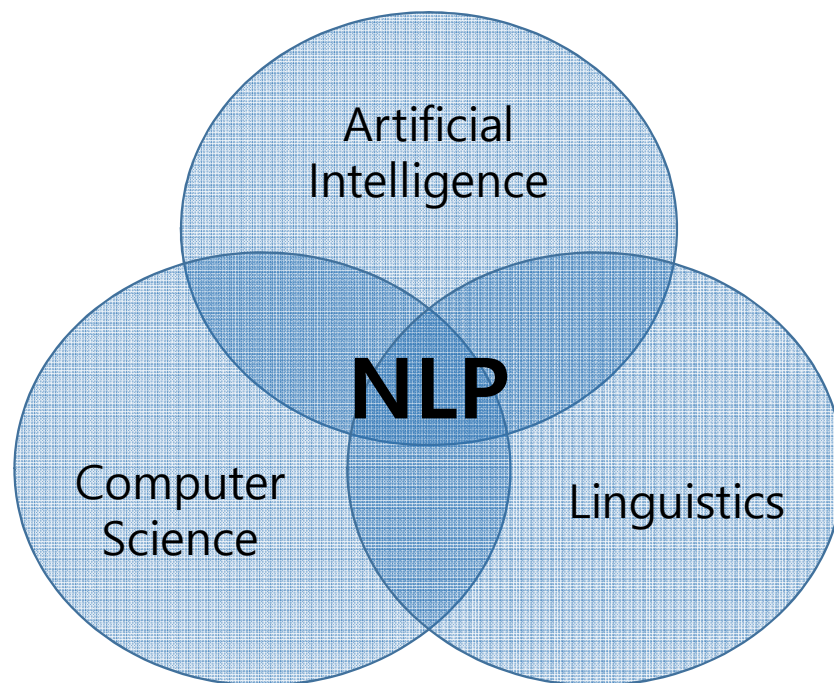
모두의연구소 Research Scientist



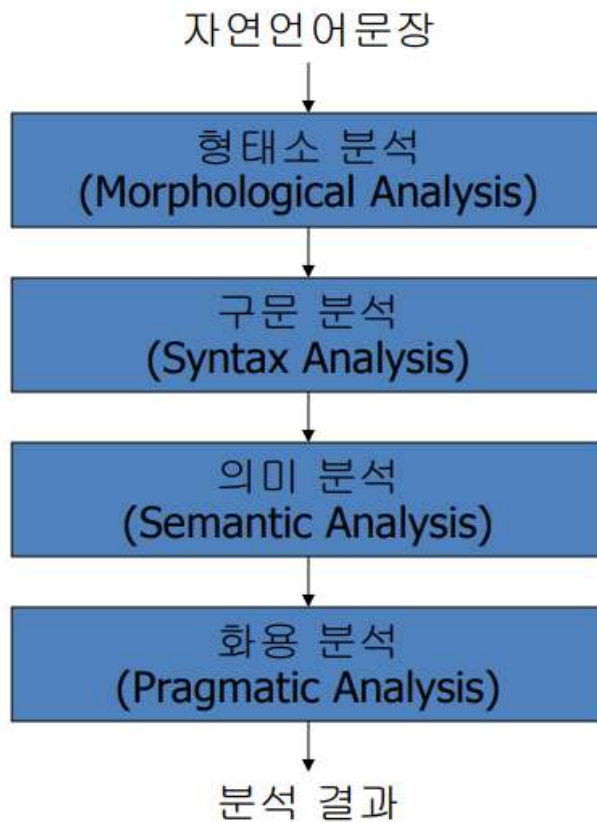
2019 모두의연구소

Natural Language Processing(NLP)

- 기계가 인간의 언어를 듣고, 이해하고, 말할 수 있게 하기 위한 노력(Since 1945, ENIAC의 탄생과 함께)



전통적인 자연어 분석단계

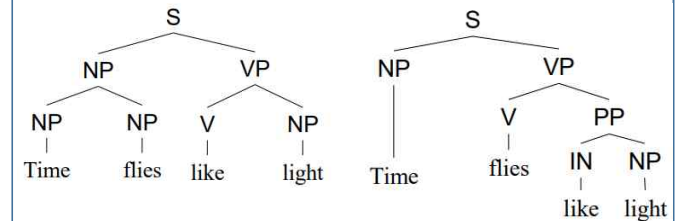


형태소 분석과 모호성

"감기는"의 분석 결과

- 감기(명사:cold) + 는(조사)
- 감(동사 어간) + 기(명사화 어미) + 는 (조사)
- 감(동사 어간) + 기는(어미)

구문 분석과 모호성



의미 분석과 모호성

의미분석과 중의성의 문제

- 말이 달린다.
 - 말이 많다.
- Horse? Speech?

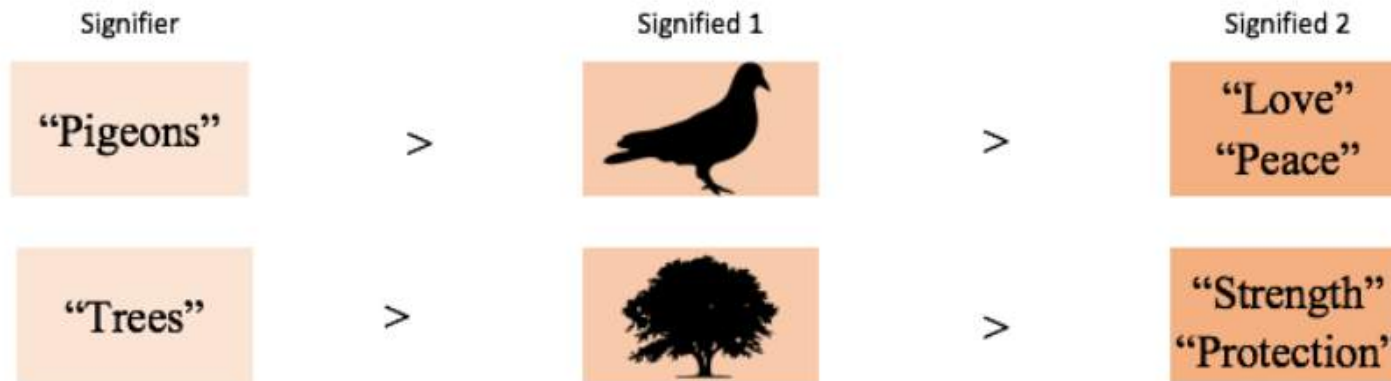
화용 분석과 상황인지

설명? 질문? 명령?

Would you mind opening the window?

NLP의 어려움

- (소쉬르) 심볼과 그 의미 사이에 필연적인 연관 관계가 존재하지 않는다



머신은 기호와 기표/기의의 연결관계에 대한 (경험적) 지식을 가지고 있지 않다.
자연어처리의 어려움인 [맥락정보의 부재에 따른 모호성의 문제]를 어떻게 해결할 것인가?

NLP as a Sequence Prediction

- Time Series Forecasting : Numeric List 의 Next step Prediction
- Sequence Prediction : Symbol List 의 Next step Prediction
- 나는 밥을 _____. 빈칸에 알맞은 말은?
- We need Representation Learning!!
 - 숫자(0, 1)밖에 이해하지 못하는 컴퓨터에게 symbol의 sequence로 이루어진 Language를 어떻게 표현해 주어야 할까?

Language Models

- Probabilistic Language Models (PLM)
 - 딥러닝 발전 이전에 시도되었던 통계적 언어 모델링 기법들
 - 언어를 확률로 표현해 주자!!
 - Uni-gram, Bi-gram, N-gram
- Neural Network Language Model (NNLM)

언어모델을 확률적으로 표현한다는 점에서 위 두 방법은 본질적으로 차이가 없다.
그러나, '맥락'을 어떻게 표현하느냐 하는 점에서 큰 차이가 있다.

Probabilistic Language Model (1)

- 확률론적 언어 모형(Probabilistic Language Model)은 m 개의 단어 w_1, w_2, \dots, w_m 열(word sequence)이 주어졌을 때 문장으로써 성립될 확률 $P(w_1, w_2, \dots, w_m)$ 을 출력함으로써 이 단어 열이 실제로 현실에서 사용될 수 있는 문장(sentence)인지를 판별하는 모형이다.
 - $p(\text{'나는', '밥을', '먹었다'}) > p(\text{'밥은', '나를', '먹었다'})$
 - 이 확률은 코퍼스 데이터 상의 word sequence 출현 빈도 통계를 통해 얻어진다.



Probabilistic Language Model (2)

- Uni-gram
 - 모든 단어의 활용이 완전히 서로 독립이라고 가정한다.

$$P(w_1, w_2, \dots, w_m) = \prod_{i=1}^m P(w_i)$$

Uni-gram은 단어들 사이의 관계(인접, 호응)를 완전 무시하므로, 완전 무맥락적 언어모델이다.

- Bi-gram
 - 단어의 활용이 바로 전 단어에만 의존한다

$$P(w_1, w_2, \dots, w_m) = P(w_1) \prod_{i=2}^m P(w_i | w_{i-1})$$

Bi-gram은 단어 출현 확률을 직전 단어 출현에 대한 조건부 확률로 표현한다. 직전 단어가 다음 단어를 해석하는 **맥락**이 된다.

- N-gram
 - 단어의 활용이 이전 N-1개의 단어에 의존한다.

$$P(w_1, w_2, \dots, w_m) = P(w_1) \prod_{i=n}^m P(w_i | w_{i-1}, \dots, w_{i-n})$$

N-gram의 N이 커질 수록 더 큰 범위의 맥락을 볼 수 있다. 그러나 계산복잡도가 지수적으로 증가한다.



Curse of Dimensionality

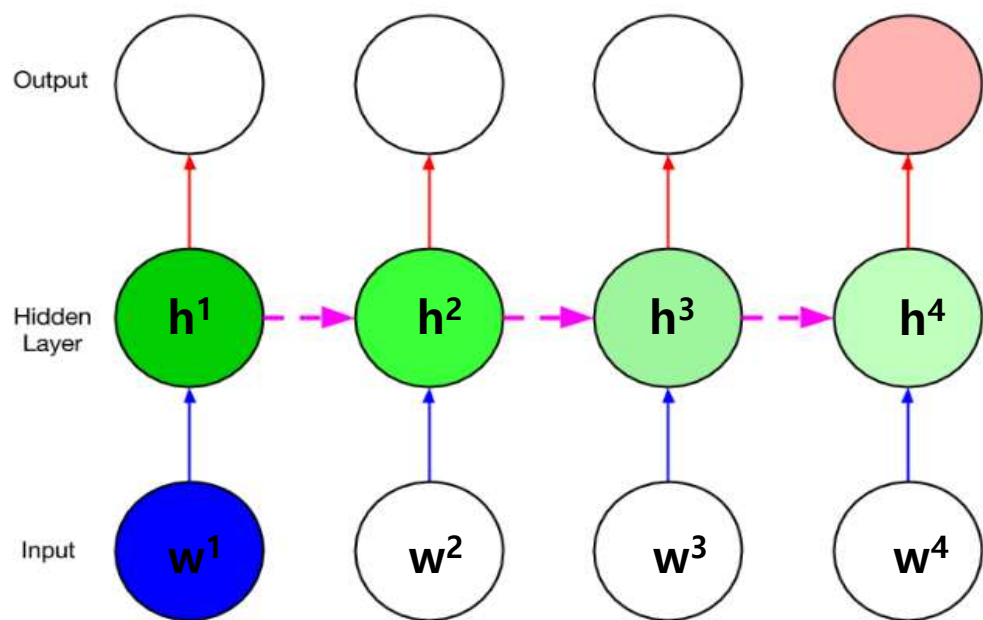
- 사전에 단어가 100,000개 있다면,
- N-gram의 N이 길수록 더 긴 맥락을 고려할 수 있으니 N=10이라고 둔다면,
- 10-gram 확률을 표현하기 위해서 필요한 파라미터 개수

$$100\,000^{10} - 1 = 10^{50} - 1$$



NLM as a Sequence Prediction

- Neural Network(특히 RNN)을 사용하면 어떨까?



$$P(w_1, w_2, \dots, w_m) = P(w_1) \prod_{i=2}^m P(w_i | w_{i-1}, \dots, w_{i-n})$$

만약 N-gram의 맥락조건 부분을 RNN의 **hidden state**로 대체한다면 어떨까?

[질문거리]
그런데, word를 RNN의 input
으로 어떻게 표현
(representation)해야 할까요?

다음 시간으로
이어집니다.



Language Model로 할 수 있는 것들

- **쉬운 문제들(Easy)**

- 철자 검사(Spell Checking)
- 키워드 검색(Keyword Search)
- 동의어 찾기(Finding Synonyms)

- **중간 난이도 문제들(Medium)**

- 웹사이트, 문서 등으로부터 정보 추출(Parsing information from websites, documents, etc.)

- **어려운 문제들(Hard)**

- 기계 번역(Machine Translation) -e.g. 중국어를 영어로 번역
- 의미 분석(Semantic Analysis) -질문의 의미가 무엇인가?
- 동일 지시어(Coreference) -e.g. 문서에서 "he"나 "it"이 지시하는 바가 무엇인가?
- 질의 응답(Question Answering) -e.g. 제퍼디 퀴즈쇼 질문에 답하기



NLP의 4단계

