CUAI Standford Univ. CS224n 스터디(NLP)

Lecture 1 – Introduction and Word Vectors

발표자 : 조민지

목차

1. Human language and word meanings

인간의 가장 강력한 무기 = 인간의 언어 (→ 네트워킹을 가능하게 함)

→ '글쓰기'의 발명 → 지식을 전파할 수 있게 됨

하지만 오늘날 주고받는 데이터의 크기에 비해 인간의 언어는 느림

→ 언어가 이토록 강력한 도구가 될 수 있는 이유는 압축성에 있음

우리가 대화를 주고받을 때 상대방이 배경지식을 가지고 있음을 전제함

→ 이미지 데이터에 비해 적은 용량으로도 정보를 전달할 수 있음

1. Human language and word meanings

'meaning'이란 무엇인가?

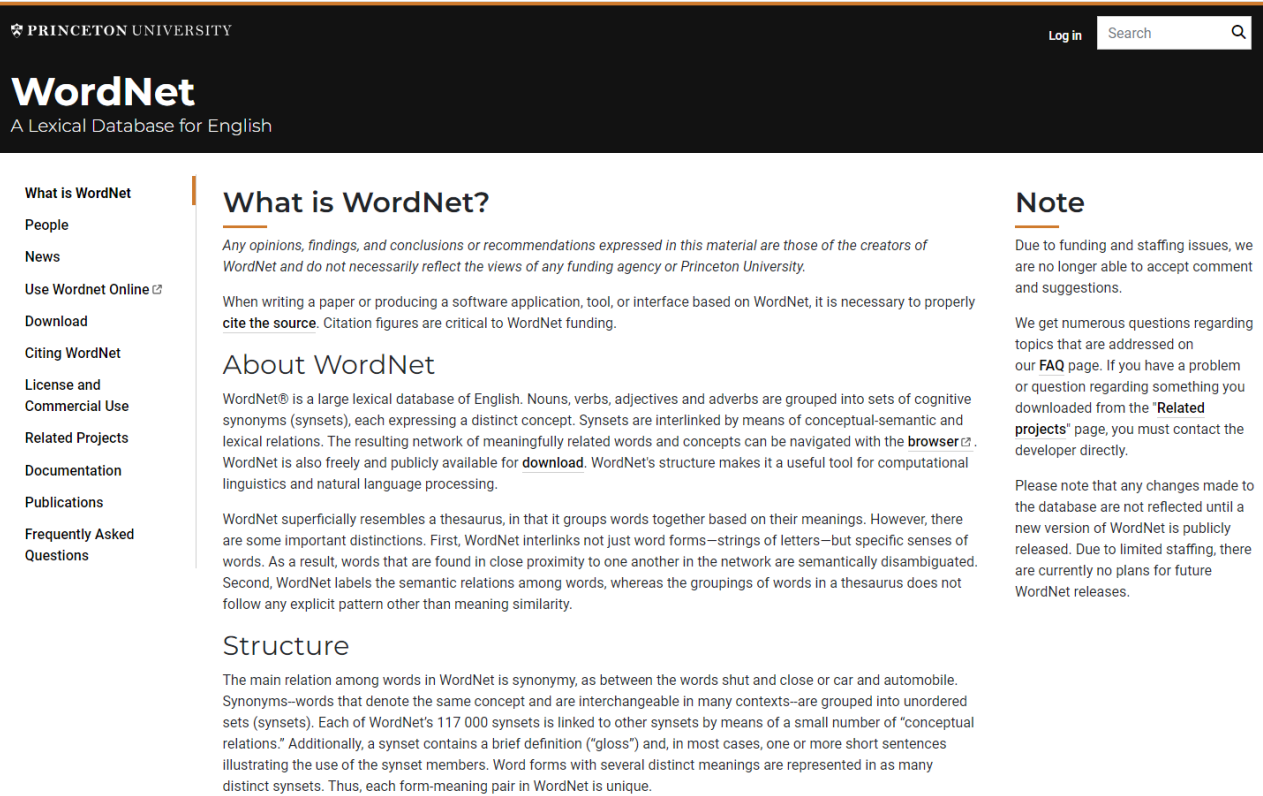
- **사전적 의미** : idea

- 'meaning'에 대한 **언어학적 해석** :

  signifier(symbol) ⇔ signified(idea of thing) : denotational semantics 표시적 의미론

- **해결책** : 시소러스 이용 (e.g., WordNet)

1. Human language and word meanings



PRINCETON UNIVERSITY

Log in Search

WordNet

A Lexical Database for English

- What is WordNet
- People
- News
- Use Wordnet Online
- Download
- Citing WordNet
- License and Commercial Use
- Related Projects
- Documentation
- Publications
- Frequently Asked Questions

What is WordNet?

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the creators of WordNet and do not necessarily reflect the views of any funding agency or Princeton University.

When writing a paper or producing a software application, tool, or interface based on WordNet, it is necessary to properly **cite the source**. Citation figures are critical to WordNet funding.

About WordNet

WordNet® is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. The resulting network of meaningfully related words and concepts can be navigated with the **browser**. WordNet is also freely and publicly available for **download**. WordNet's structure makes it a useful tool for computational linguistics and natural language processing.

WordNet superficially resembles a thesaurus, in that it groups words together based on their meanings. However, there are some important distinctions. First, WordNet interlinks not just word forms—strings of letters—but specific senses of words. As a result, words that are found in close proximity to one another in the network are semantically disambiguated. Second, WordNet labels the semantic relations among words, whereas the groupings of words in a thesaurus does not follow any explicit pattern other than meaning similarity.

Structure

The main relation among words in WordNet is synonymy, as between the words shut and close or car and automobile. Synonyms--words that denote the same concept and are interchangeable in many contexts--are grouped into unordered sets (synsets). Each of WordNet's 117 000 synsets is linked to other synsets by means of a small number of "conceptual relations." Additionally, a synset contains a brief definition ("gloss") and, in most cases, one or more short sentences illustrating the use of the synset members. Word forms with several distinct meanings are represented in as many distinct synsets. Thus, each form-meaning pair in WordNet is unique.

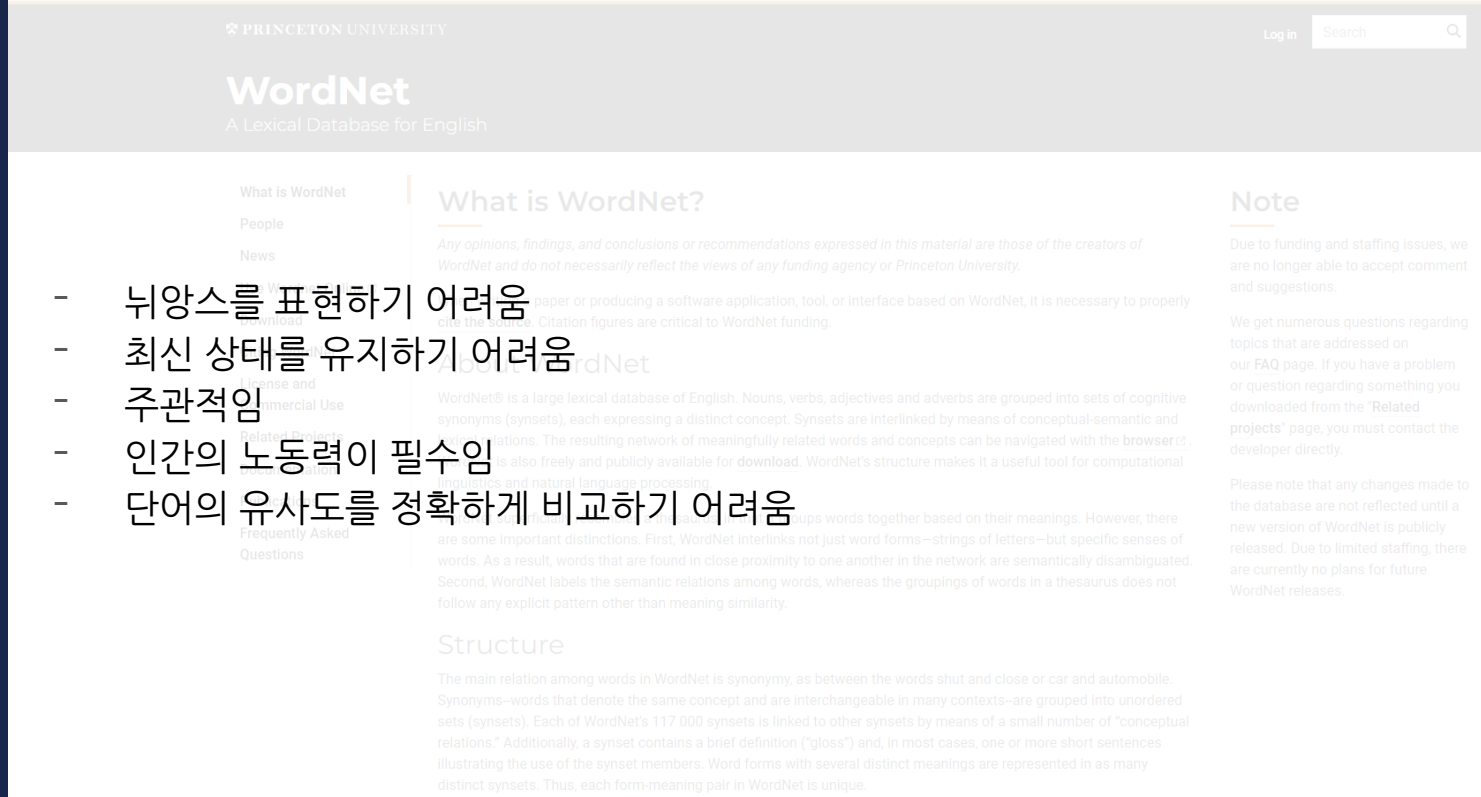
Note

Due to funding and staffing issues, we are no longer able to accept comment and suggestions.

We get numerous questions regarding topics that are addressed on our **FAQ** page. If you have a problem or question regarding something you downloaded from the "**Related projects**" page, you must contact the developer directly.

Please note that any changes made to the database are not reflected until a new version of WordNet is publicly released. Due to limited staffing, there are currently no plans for future WordNet releases.

1. Human language and word meanings



- 뉘앙스를 표현하기 어려움
- 최신 상태를 유지하기 어려움
- 주관적임
- 인간의 노동력이 필수임
- 단어의 유사도를 정확하게 비교하기 어려움

1. Human language and word meanings

단어를 Discrete Symbol로 표현 (~2012년) : one-hot vectors

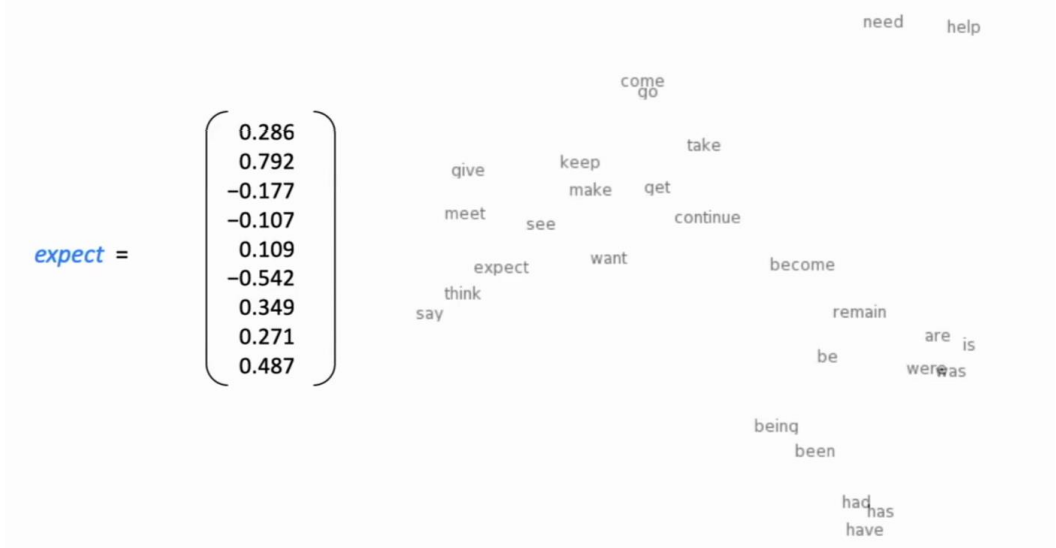
Motel = [0 0 0 0 0 0 0 0 0 0 1 0 0 0 0]
Hotel = [0 0 0 0 0 0 0 1 0 0 0 0 0 0 0]

- 단어의 수만큼 벡터의 차원이 높아짐
- 유사한 의미의 두 단어들의 벡터가 orthogonal할 수 있음

1. Human language and word meanings

단어를 **문맥**에 따라 표현 (2013년 ~) : Distributional semantics

단어의 의미를 근처에 빈번하게 나타나는 단어에 따라 받아들임



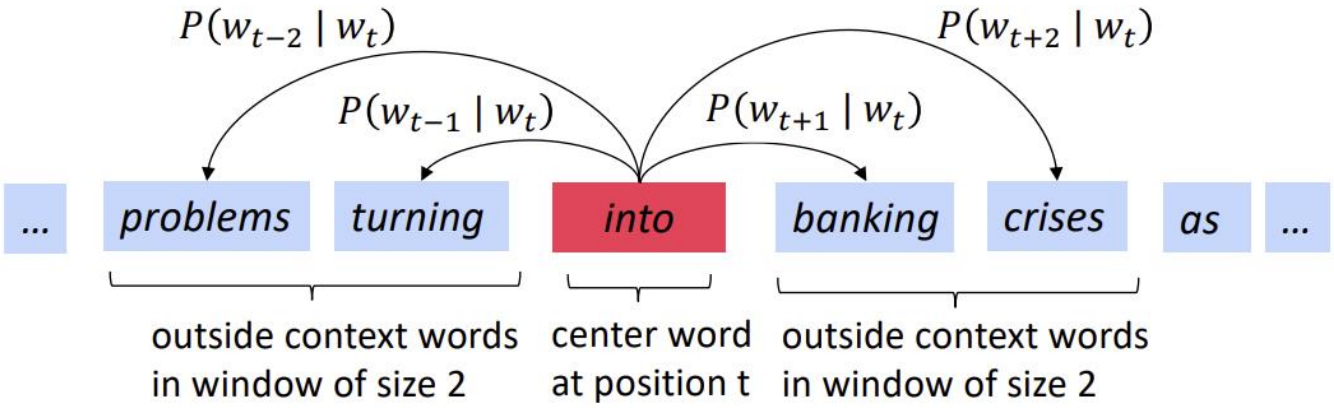
2. Word2vec Introduction

Word2vec

word Vector를 학습하기 위한 프레임워크 (Mikolov et al. 2013)

- 대규모의 말뭉치(corpus)를 갖고 있음
- 모든 단어가 고정된 어휘(vocabulary) 안에 존재하고 벡터로 표현됨
- Center word ‘c’와 context words ‘o’를 갖는 위치 ‘t’를 지남
- C에 대한 o의 확률을 계산하기 위해 word vectors의 유사도를 이용함
- 확률을 최대화하기 위해 words vector를 조정함

2. Word2vec Introduction



2. Word2vec Introduction

$$L(\theta) = \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \neq 0}} P\left(w_{t+j} \mid w_t; \theta\right)$$

$$J(\theta) = -\frac{1}{T} \log L(\theta) = -\frac{1}{T} \sum_{t=1}^{T} \sum_{\substack{-m \le j \le m \\ j \neq 0}} \log P\left(w_{t+j} \mid w_t; \theta\right)$$

2. Word2vec Introduction

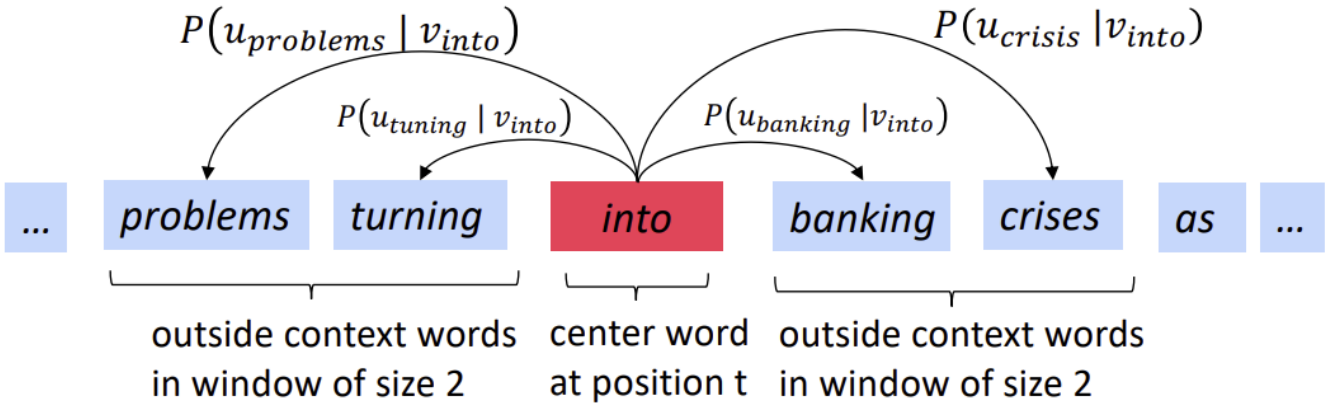
$$J(\theta) = -\frac{1}{T}\sum_{t=1}^{T}\sum_{\substack{-m\le j\le m \\ j\neq 0}} \log P(w_{t+j} \mid w_t;\theta)$$ ← minimize the objective funtion

$$P(w_{t+j}|w_t;\theta)$$

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w\in V}\exp(u_w^T v_c)}$$

v_w : w는 centor word
u_w : w는 context word

2. Word2vec Introduction

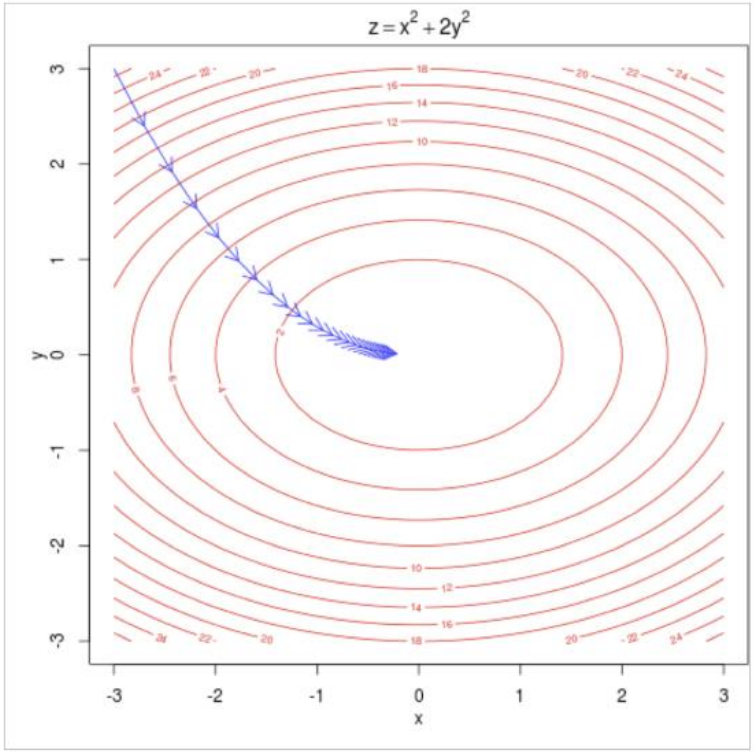


2. Word2vec Introduction

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w \in V} \exp(u_w^T v_c)}$$

$$Softmax(x_i) = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} = p_i$$

3. Word2vec objective function gradient



3. Word2vec objective function gradient

$$\max J'(\theta) = \prod_{t=1}^{T} \prod_{\substack{-m \le j \le m \\ j \neq 0}} P(w'_{t+j} | w_t; \theta)$$

$$\min J(\theta) = -\frac{1}{T} \sum_{t=1}^{T} \sum_{\substack{-m \le j \le m \\ j \neq 0}} \log P(w'_{t+j} | w_t)$$

$$P(o|c) = \frac{\exp(u_o^T v_c)}{\sum_{w=1} \exp(u_w^T v_c)}$$

3. Word2vec objective function gradient

$$\frac{\partial}{\partial v_c}\log P(o|c) = \frac{\partial}{\partial v_c}\log\frac{\exp(u_o^T v_c)}{\sum_{w=1}^{v}\exp(u_o^T v_c)}$$

$$= \frac{\partial}{\partial v_c}\log\exp(u_o^T v_c) - \frac{\partial}{\partial v_c}\log\sum_{w=1}^{v}\exp(u_o^T v_c)$$

$$= \frac{\partial}{\partial v_c}\ u_o^T v_c - \frac{\partial}{\partial v_c}\log\sum_{w=1}^{v}\exp(u_o^T v_c)$$

$$= u_o - \sum_{X=1}^{v} P(X|c)\cdot u_X$$

3. Word2vec objective function gradient

$$\frac{\partial}{\partial v_c}\log\sum_{w=1}^{v}\exp(u_o^T v_c) \quad \rightarrow \text{Chain Rule}$$

$$= \frac{1}{\sum_{w=1}^{v}\exp(u_o^T v_c)}\cdot\frac{\partial}{\partial v_c}\sum_{X=1}^{v}\exp(u_X^T v_c)$$

$$= \frac{1}{\sum_{w=1}^{v}\exp(u_o^T v_c)}\cdot\sum_{X=1}^{v}\frac{\partial}{\partial v_c}\exp(u_X^T v_c) \quad \rightarrow \text{Chain Rule}$$

$$= \frac{1}{\sum_{w=1}^{v}\exp(u_o^T v_c)}\cdot\sum_{X=1}^{v}\exp(u_X^T v_c)\cdot\frac{\partial}{\partial v_c}u_X^T v_c$$

$$= \frac{1}{\sum_{w=1}^{v}\exp(u_o^T v_c)}\cdot\sum_{X=1}^{v}\exp(u_X^T v_c)\cdot u_X$$

$$= \sum_{X=1}^{v}\frac{\exp(u_X^T v_c)}{\sum_{w=1}^{v}\exp(u_o^T v_c)}\cdot u_X = \sum_{X=1}^{v}P(X|c)\cdot u_X$$