Lecture 1 preview

- **Language**
  - Language is uncertain evolved system of communication but somehow we have enough agreed meaning
  - Using language not only for the information functions but also for the social functions

  - human convey knowledge through human language
  - Human beings can communicate with each other therefore work much more effective in teams, this made humans invincible
  - Humans invented writing : knowledge could sent spatially and temporally
  - Compared to computer network, human language is very slow network
  - human language can carry huge amount of compressed information

- **Meaning**
  - definition of meaning : denotational semantics

- **How do we have usable meaning in a computer?**
  1. WordNet : a thesaurus containing lists of synonym, hypernym sets
     - problems :
       - missing nuance
       - missing new words
       - humans understand similarities and relations btw the meaning of words, but WordNet cannot. Because it has fixed sets.

  2. one-hot vectors

- problems :
    - infinite number of words(derivational morphology)→ infinite dimentions
    - all vectors are orthogonal : cannot inform about relationships or meaning of words

3. distributional semantics
    - a word’s meaning is given by the words that frequently appear close-by
    - context words represent the word
    - build a dense vector for each word, plus all of the numbers are non-zero
    - One-hot 벡터보다 작은 차원 내에서 표현이 가능하다
    - Vector space내에 단어를 배치한다. n차원 벡터공간의 표현이 어렵기 때문에 2차원 벡터공간으로 사영되어 표현되지만(projection may misrepresent what’s in the original space), 단어 간의 유사성을 볼 수 있다.

- **Word2vec overview**
    - Framework for learning word vectors
    - idea :
        1. We have large body(corpus) of text
        2. Every word in a fixed vocabulary is represented by a vector
        3. Random vectors 로 시작, big iterative algorithm where we go through each position in the text
        4. Meaning of the word=context : we want Representation of the word in the middle to be able to predict the words that are around it → 반복을 통해 단어의 position을 조정해가며 맞는 위치를 찾는다

- **Word2vec : objective function**
    - likelihood
    - objective function를 최소화하는 방향으로 vector representation 찾아내야 한다.

- minus : minimization rather than maximize
- 1/T : keeps the scale of things, not dependent of the size of the corpus
- Log(important) : multiplication을 sums of probability 로 바꿔준다

How to calculate probability? → Use 2 vectors per word W : when W is a context word, when W is a center word

- **Word2vec : prediction function**
  - dot product : similarity btw o and c. If similar, the dot product will be big.
  - Exp : makes every number positive
  - Sum up the quantity, divide the dot product. → normalize over entire vocabulary to give probability distribution
  - Softmax distribution (입력받은 값을 출력으로 0~1사이의 값으로 모두 정규화하며 출력 값들의 총합은 항상 1이 되는 특성을 가진 함수) → 최댓값의 soft한 근사치를 찾아낸다

- Training a model by optimizing parameters
  - To train a model, we adjust parameters to minimize a loss : compute all vector gradients