Word Embeddings

Outline

- NLP Intro
- Word Representations and word embeddings
- Word2Vec model
- Glove model

NLP - Natural Language Processing

NLP is the field that includes : understanding, processing, analyzing and generating natural languages.

Applications in NLP :-

- Translation
- Information Extraction
- Summarization
- Parsing
- Q/A
- Sentiment Analysis and much more

NLP challenges

- Polysemy
- Syntactic ambiguity
- Variability
- Coreference resolution

Word Representation

We can represent objects in different hierarchy levels - Documents, Sentences, Phrases, Words

Goal : We want the representation to be interpretable and easy-to-use.

Vector representation meets those requirements.

The Distributional Hypothesis

Words which occur in the same contexts tends to have similar meanings. For example :

- Soundtrack, lyrics, sung, duet : Song
- Cucumber, sauce, pizza, ketchup : Tomato

Vector Representation

We can define a word by a vector of counts over contexts.

	song	cucumber	meal	black
tomato	**0**	**6**	**5**	**0**
book	2	0	2	3
pizza	**0**	**2**	**4**	**1**

- Each word is associated with a vector of dimension mod(V) which is the size of vocabulary.
- We expect the similar words to have similar vectors.
- Given the vectors of two words, we can determine their similarity.

From Sparse to Dense

These vectors are :

- Huge each of dimension mod(V)
- Sparse since most of the entries are 0

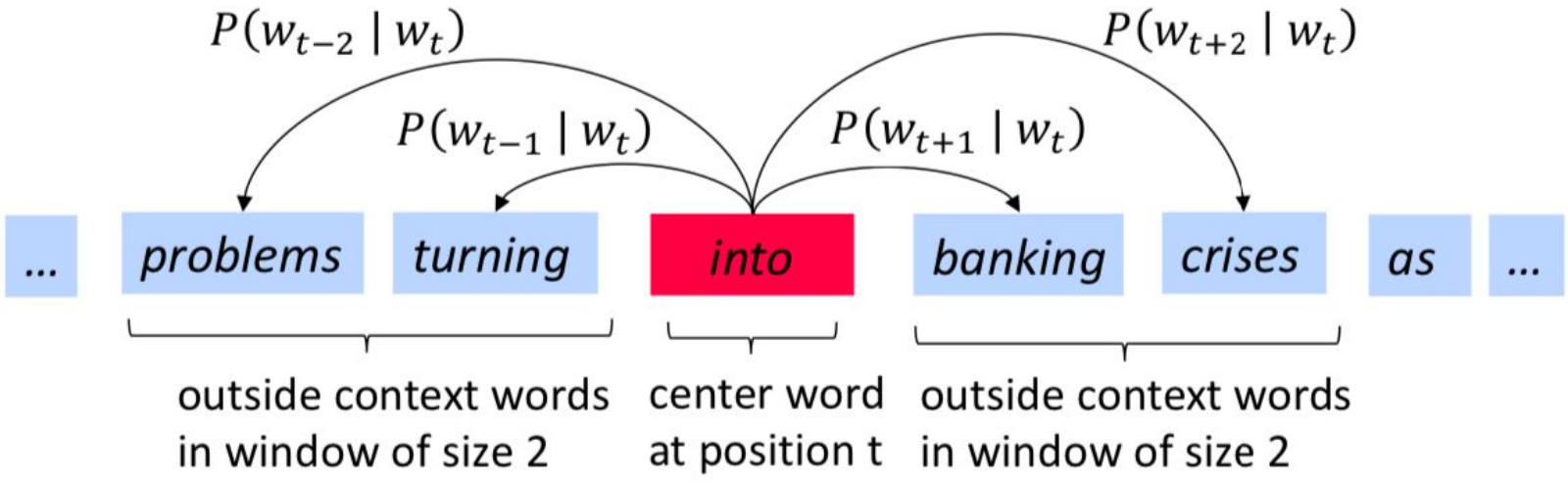
So, we want our vectors to be small and dense :

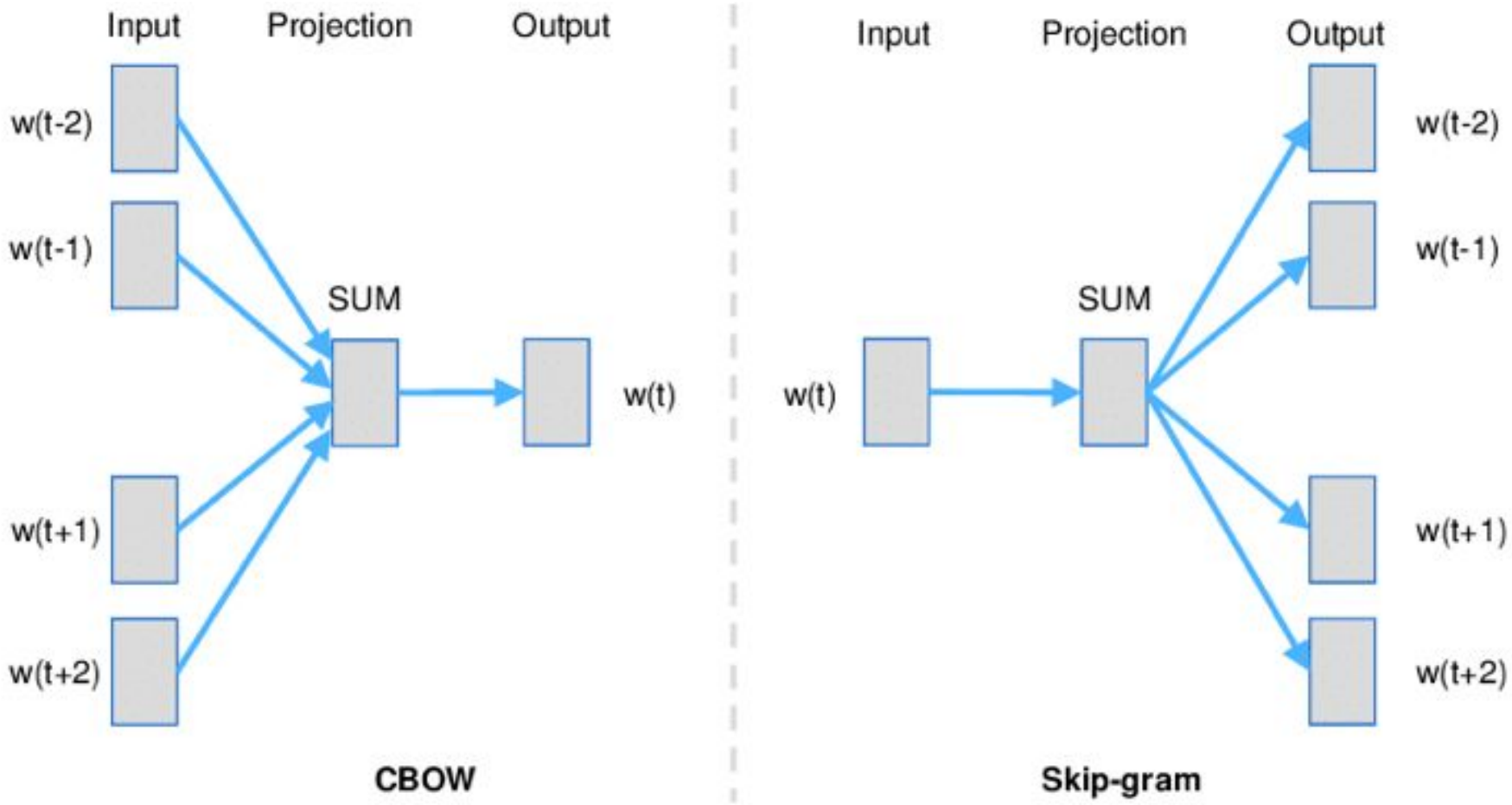
- Use reduction algorithm like SVD
- Low dimensional word-vectors directly : word embeddings

Word2Vec

It is a framework for learning word vectors.Main idea :

- Every word in the fixed vocabulary is represented by a vector.
- Go through each position t in the text, which has the center word: c and context words: o
- Use the similarity of the word vectors for o and c to calculate the probability of o given c or vice-versa
- Keep adjusting word vector such that the probability is maximum


$P(w_{t-2} \mid w_t)$
$P(w_{t-1} \mid w_t)$
$P(w_{t+1} \mid w_t)$
$P(w_{t+2} \mid w_t)$
...
problems
turning
into
banking
crises
as
...
outside context words in window of size 2
center word at position t
outside context words in window of size 2


Input
Projection
Output
w(t-2)
w(t-1)
SUM
w(t)
w(t+1)
w(t+2)
CBOW
Input
Projection
Output
w(t)
SUM
w(t-2)
w(t-1)
w(t+1)
w(t+2)
Skip-gram

Glove

GloVe is an unsupervised learning algorithm for obtaining vector representations for words. Training is performed on aggregated global word-word co-occurrence statistics from a corpus, and the resulting representations showcase interesting linear substructures of the word vector space.