IN5550: Neural Methods in Natural Language Processing
Sub-lecture 11.1
Problems of static word embeddings

Andrey Kutuzov

University of Oslo

31 March 2022



Contents



1 Brief Recap

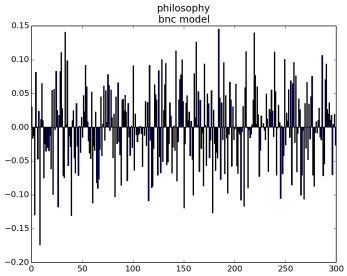
2 Problems of static word embeddings

Brief Recap

Word embeddings

- Distributional models are based on distributions of word co-occurrences in large training corpora;
- they represent lexical meanings as dense vectors (embeddings);
- the representations are also distributed: meaning is expressed via values of multiple vector entries;
- particular vector entries (features) are not directly related to any particular semantic 'properties', and thus not directly interpretable;
- words occurring in similar contexts have similar vectors.

Important: each word is associated with exactly one dense vector.
Hence, such models are sometimes called 'static embeddings'.



Brief Recap



Word (or subword) embeddings are often used as an input to neural network models:

Brief Recap



Word (or subword) embeddings are often used as an input to neural network models:

- feed-forward networks,

Brief Recap



Word (or subword) embeddings are often used as an input to neural network models:

- feed-forward networks,
- convolutional networks,

Brief Recap



Word (or subword) embeddings are often used as an input to neural network models:

- feed-forward networks,
- convolutional networks,
- recurrent networks: LSTMs, GRUs, etc (Obligatory 2)

Brief Recap



Word (or subword) embeddings are often used as an input to neural network models:

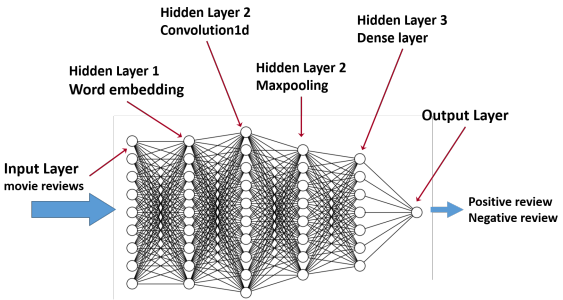
- feed-forward networks,
- convolutional networks,
- recurrent networks: LSTMs, GRUs, etc (Obligatory 2)
- transformers (Obligatory 3).

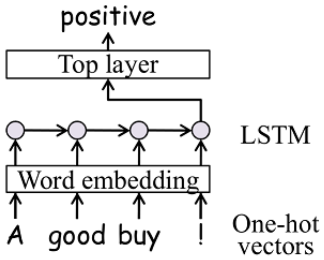
Brief Recap

Word (or subword) embeddings are often used as an input to neural network models:

- feed-forward networks,
- convolutional networks,
- recurrent networks: LSTMs, GRUs, etc (Obligatory 2)
- transformers (Obligatory 3).

Embeddings themselves can be updated at the training time along with the rest of the network weights or 'frozen' (protected from updating).





Contents



1 Brief Recap

2 Problems of static word embeddings

Problems of static word embeddings

Meaning is meaningful only in context

English sentences with the word 'bank' in two different senses:

1. '*She was enjoying her walk down the quiet country lane towards the river bank.*' (**sense 0**)
2. '*The bank obtained its customer payment or reimbursement and released the goods to the customer.*' (**sense 1**)

Problems of static word embeddings

Meaning is meaningful only in context

English sentences with the word 'bank' in two different senses:

1. '*She was enjoying her walk down the quiet country lane towards the river bank.*' (**sense 0**)
2. '*The bank obtained its customer payment or reimbursement and released the goods to the customer.*' (**sense 1**)

Even most perfect 'static' embeddings will always yield the same vector for 'bank' . But in fact the meaning is different!

Problems of static word embeddings

Meaning is meaningful only in context

English sentences with the word 'bank' in two different senses:

1. '*She was enjoying her walk down the quiet country lane towards the river bank.*' (**sense 0**)
2. '*The bank obtained its customer payment or reimbursement and released the goods to the customer.*' (**sense 1**)

Even most perfect 'static' embeddings will always yield the same vector for 'bank' . But in fact the meaning is different!

What can be done?

Problems of static word embeddings

One can represent 'bank' as the average embedding of all the context words [Schütze, 1998].
But:

Problems of static word embeddings

One can represent 'bank' as the average embedding of all the context words [Schütze, 1998].
But:

1. Context words themselves can be ambiguous.

Problems of static word embeddings

One can represent 'bank' as the average embedding of all the context words [Schütze, 1998].
But:

1. Context words themselves can be ambiguous.
2. Their contextual senses will be lost.

Problems of static word embeddings

One can represent 'bank' as the average embedding of all the context words [Schütze, 1998].
But:

1. Context words themselves can be ambiguous.
2. Their contextual senses will be lost.
3. In this 'bag of embeddings', word order information is also entirely lost.



So, what to do?

References I



Schütze, H. (1998).
Automatic word sense discrimination.
Computational Linguistics, 24(1):97–123.