# CS 6501 Natural Language Processing

## Text Representation Learning

---

Yangfeng Ji

November 19, 2018

Department of Computer Science
University of Virginia

# Overview

1. Principle of Compositionality
2. Modeling Composition Functions
3. Generalized Distributional Hypothesis

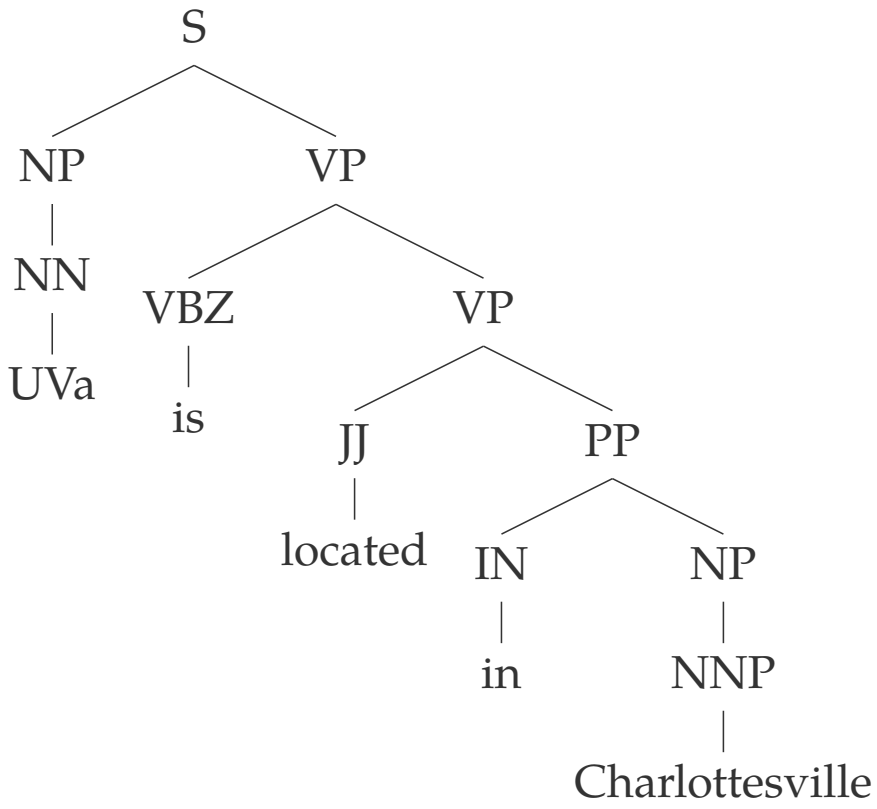
# Principle of Compositionality

# Principle of Compositionality

**Principle [Partee, 2007]**

The meaning of a whole is a function of the meanings of the parts and of the way they are syntactically combined.

# Example

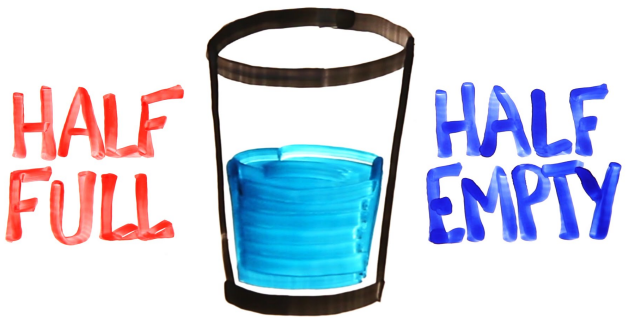

S
NP
VP
NN
VBZ
VP
UVa
is
JJ
PP
located
IN
NP
in
NNP
Charlottesville

# Challenge of Composition Operations

- `empty`
- `full`

# Challenge of Composition Operations

- `empty`
- `full`
- `half empty`
- `half full`



# Challenge of Composition Operations

Example

- `good`
- `not good`
- `not good at all`

# Challenge of Composition Operations

Example

- `good`
- `not good`
- `not good at all`
- `very good`
- `not very good`

# Modeling Composition Functions

# Linear Operations

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ are two embeddings, the composition function can be represented as

$$\boldsymbol{p} = f(\boldsymbol{u}, \boldsymbol{v}, R, K) \tag{1}$$

where $R$ is syntactic relation between $\boldsymbol{u}$ and $\boldsymbol{v}$, and $K$ denotes some additional knowledge for composition.

[Mitchell and Lapata, 2008]

# Simple Examples

Without $R$ and $K$, the composition function is simplified as $f(u, v)$ with the following special cases

- $p = u + v$
- $p = u \circ v$

[Mitchell and Lapata, 2008]

# Simple Examples

Without $R$ and $K$, the composition function is simplified as $f(u, v)$ with the following special cases

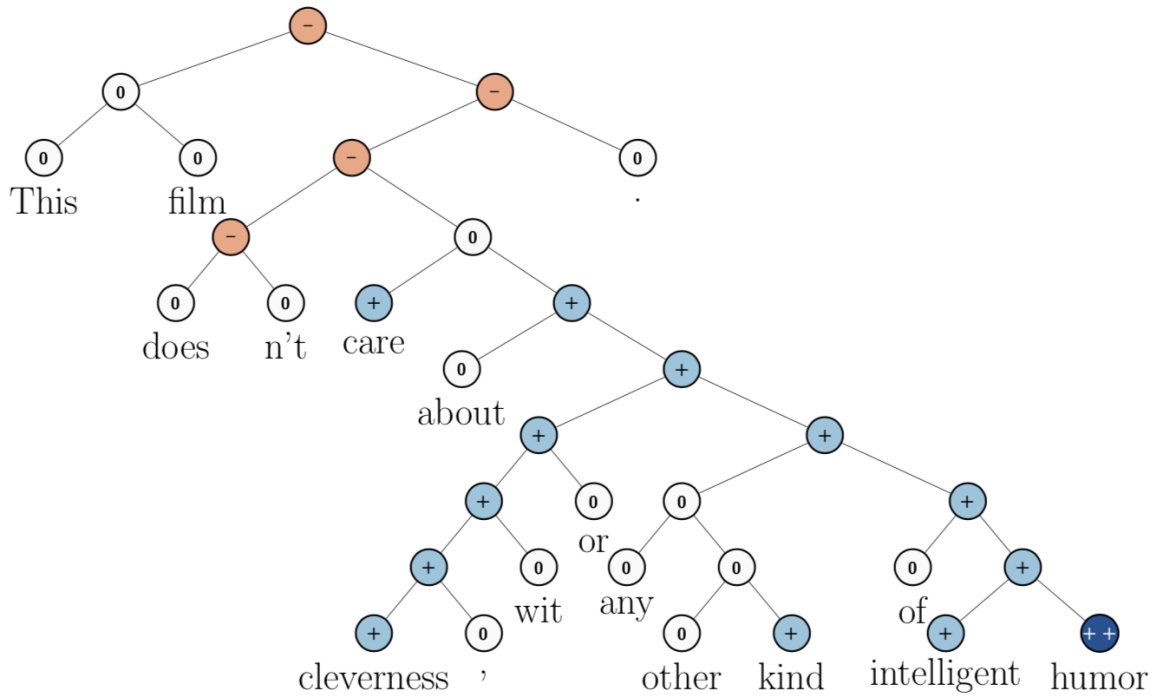
- $p = u + v$
- $p = \mathbf{A}u + \mathbf{B}v$
- $p = u \circ v$
- $p = u\mathbf{C}v$

where $\mathbf{A}$ and $\mathbf{B}$ are parameterized matrices, and $\mathbf{C}$ is a 3-D parameterized tensor

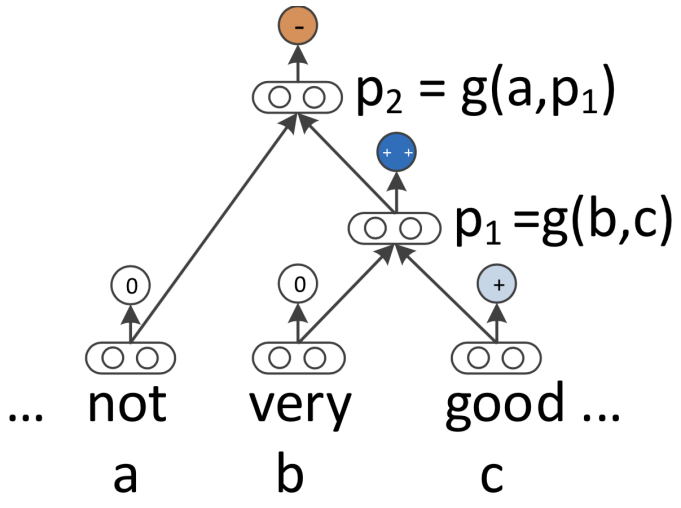
[Mitchell and Lapata, 2008]

# Composition on Sentences

This film does n't care about cleverness, wit or any other kind of intelligent humor

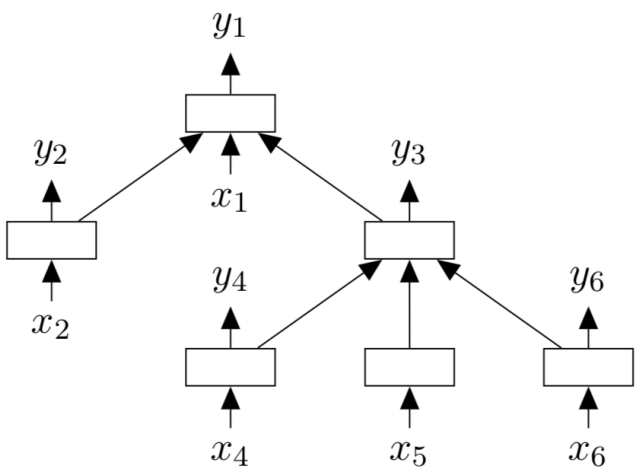


# Recursive Neural Networks




$$p_1 = f\left(W\begin{bmatrix} b \\ c \end{bmatrix}\right), p_2 = f\left(W\begin{bmatrix} a \\ p_1 \end{bmatrix}\right)$$

# Tree-LSTM




$$\tilde{h}_j = \sum_{k \in C(j)} h_k,$$

$$i_j = \sigma\left(W^{(i)} x_j + U^{(i)} \tilde{h}_j + b^{(i)}\right),$$

$$f_{jk} = \sigma\left(W^{(f)} x_j + U^{(f)} h_k + b^{(f)}\right),$$

$$o_j = \sigma\left(W^{(o)} x_j + U^{(o)} \tilde{h}_j + b^{(o)}\right),$$

$$u_j = \tanh\left(W^{(u)} x_j + U^{(u)} \tilde{h}_j + b^{(u)}\right),$$

$$c_j = i_j \odot u_j + \sum_{k \in C(j)} f_{jk} \odot c_k,$$

$$h_j = o_j \odot \tanh(c_j),$$

# Generalized Distributional Hypothesis

# Distributonal Hypothesis

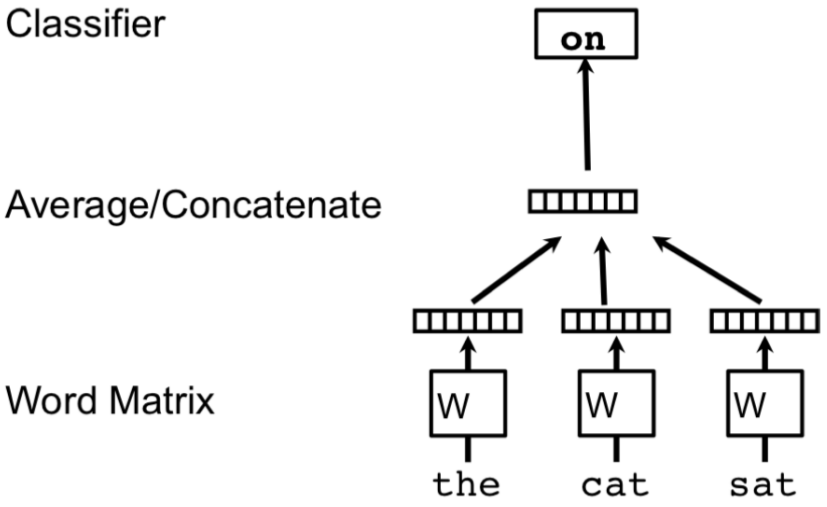
words that occur in the same contexts tend to have similar meanings

- `to have a splendid time in Rome`
- `to have a wonderful time in Rome`

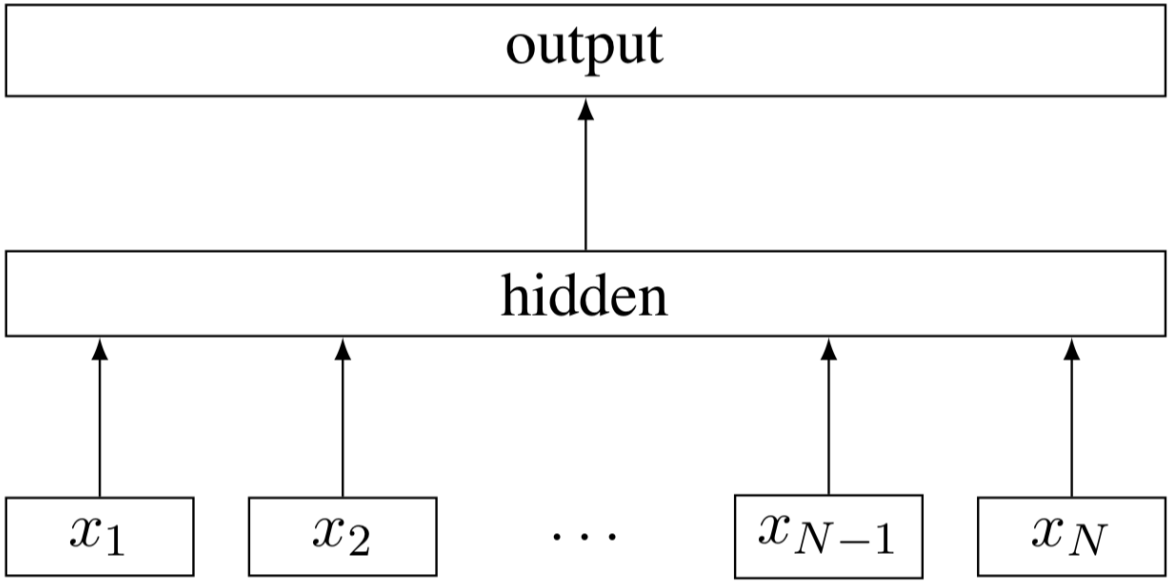
# Generalized Distributonal Hypothesis

___ that occur in the same contexts tend to have similar meanings

# Continuous BoW



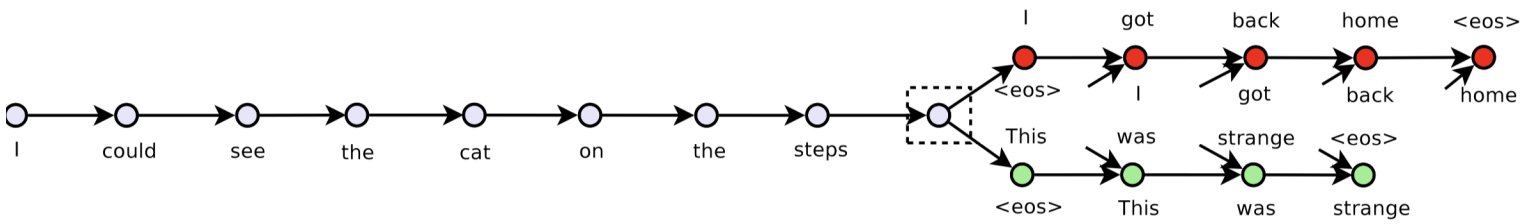
# fastText




$$-\frac{1}{N}\sum_{n=1}^{N} y_n \log(f(BAx_n)) \tag{2}$$

where $x_n$ is the normalized bag of features of the $n$-th document, $y_n$ is the label, $A$ and $B$ are the weight metrics. [Joulin et al., 2017]

# Skip-thoughts Model




[Kiros et al., 2015]

# Summary

1. Principle of Compositionality

2. Modeling Composition Functions

3. Generalized Distributional Hypothesis

# Reference


Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T. (2017).
Bag of tricks for efficient text classification.
In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 427–431. Association for Computational Linguistics.

Kiros, R., Zhu, Y., Salakhutdinov, R. R., Zemel, R., Urtasun, R., Torralba, A., and Fidler, S. (2015).
Skip-thought vectors.
In *Advances in neural information processing systems*, pages 3294–3302.

Mitchell, J. and Lapata, M. (2008).
Vector-based models of semantic composition.
*proceedings of ACL-08: HLT*, pages 236–244.

Partee, B. (2007).
Compositionality and coercion in semantics: The dynamics of adjective meaning.