GloVe: Global Vectors for Word Representation

Jeffrey Pennington, Richard Socher, Christopher D. Manning

Presented by Chris Kedzie

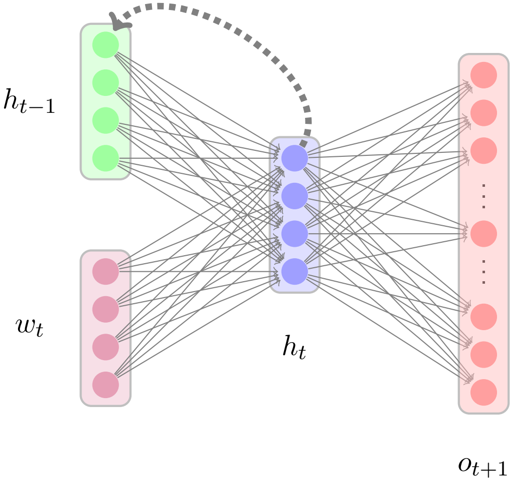
March 25, 2015

Overview

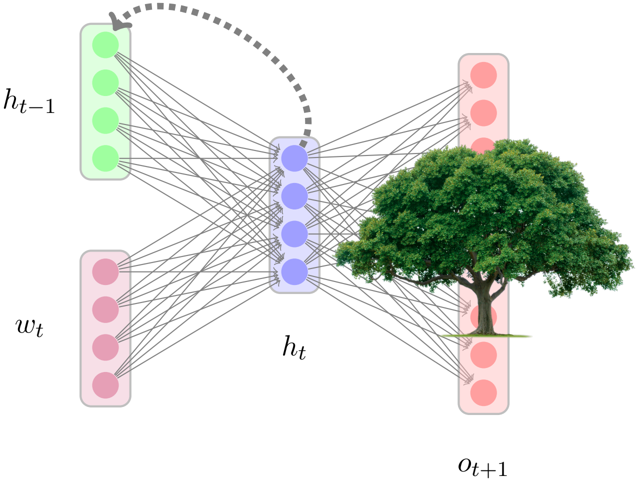
1. Introduction
2. Problem
3. GloVe Model
4. Experiments

Word Representations: A history

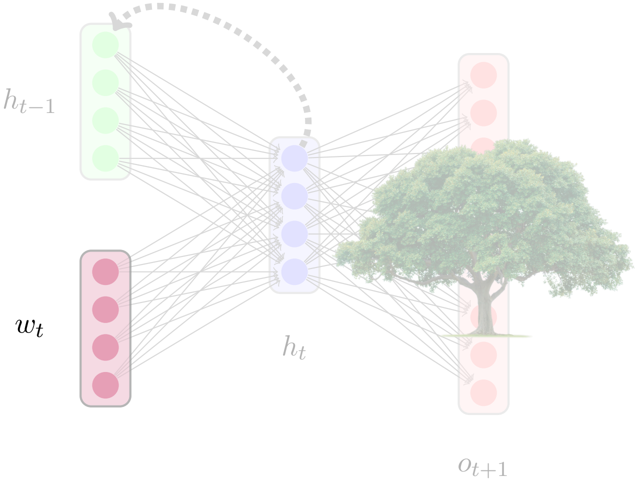
Neural Language Models – Recurrent NNLM



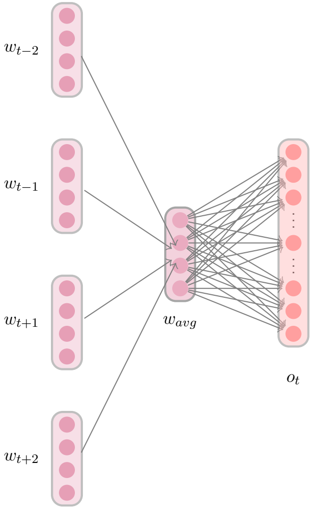
Neural Language Models – Recurrent NNLM



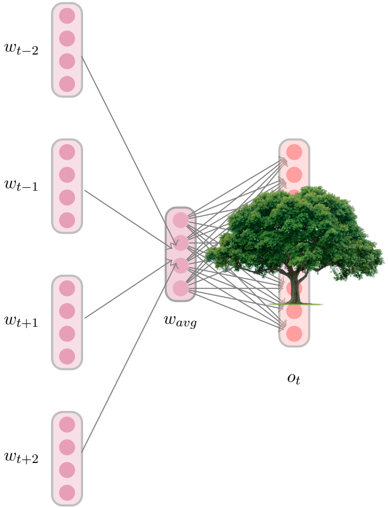
Neural Language Models – Recurrent NNLM



Neural Language Models – Continuous BOW



Neural Language Models – Continuous BOW



Linear Relationships

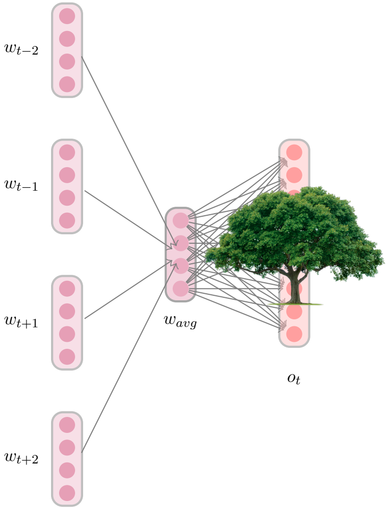
Semantic

$$w_{king} - w_{man} + w_{woman} \approx w_{queen}$$

Syntactic

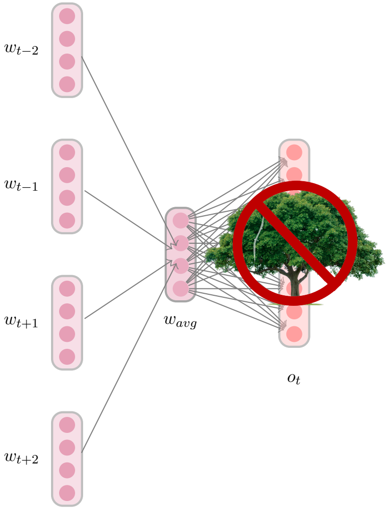
$$w_{easy} - w_{easiest} + w_{luckiest} \approx w_{lucky}$$

Noise Contrastive Estimation



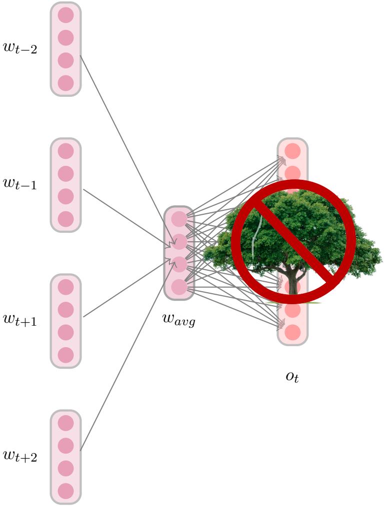
Scalable Embedding Learning

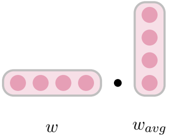
Noise Contrastive Estimation –*no more normalization required!*



Scalable Embedding Learning

Noise Contrastive Estimation –*no more normalization required!*





GloVe

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe"). Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

Local Online Optimization

Lots of time spent scanning context windows to learn a distribution for

$$P(w|the)$$

Theatre or theater is a collaborative form of fine art that uses live performers to present the experience of a real or imagined event before a live audience in a specific place. The performers may communicate this experience to the audience through combinations of gesture, speech, song, music, and dance. Elements of art and stagecraft are used to enhance the physicality, presence and immediacy of the experience. The specific place of the performance is also named by the word "theatre" as derived from the Ancient Greek (thatron, "a place for viewing"), itself from (theomai, "to see", "to watch", "to observe").
Modern Western theatre comes from large measure from ancient Greek drama, from which it borrows technical terminology, classification into genres, and many of its themes, stock characters, and plot elements. Theatre artist Patrice Pavis defines theatricality, theatrical language, stage writing, and the specificity of theatre as synonymous expressions that differentiate theatre from the other performing arts, literature, and the arts in general.
Theatre today, broadly defined, includes performances of plays and musicals, ballets, operas and various other forms.

There's got to be a better way!

Matrix Factorization Methods

e.g. SVD, COALS, etc. directly on co-occurrence matrix.

Matrix Factorization Methods

e.g. SVD, COALS, etc. directly on co-occurrence matrix.

Main drawback: frequent words like *the* and *a* have an outsized effect on the representation learning.

GloVe

GloVe Model

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix
- X_{ij} frequency of word i co-occurring with word j

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix
- X_{ij} frequency of word i co-occurring with word j
- $X_i = \sum_k^V X_{ik}$ total number of occurrences of word i in corpus

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix
- X_{ij} frequency of word i co-occurring with word j
- $X_i = \sum_k^V X_{ik}$ total number of occurrences of word i in corpus
- $P_{ij} = P(j|i) = \frac{X_{ij}}{X_i}$ a.k.a. probability of word j occurring within the context of word i

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix
- X_{ij} frequency of word i co-occurring with word j
- $X_i = \sum_k^V X_{ik}$ total number of occurrences of word i in corpus
- $P_{ij} = P(j|i) = \frac{X_{ij}}{X_i}$ a.k.a. probability of word j occurring within the context of word i
- $w \in \mathbb{R}^d$ a word embedding of dimension d

Notation!

- $X \in \mathbb{R}^{V \times V}$ word co-occurrence matrix
- X_{ij} frequency of word i co-occurring with word j
- $X_i = \sum_k^V X_{ik}$ total number of occurrences of word i in corpus
- $P_{ij} = P(j|i) = \frac{X_{ij}}{X_i}$ a.k.a. probability of word j occurring within the context of word i
- $w \in \mathbb{R}^d$ a word embedding of dimension d
- $\tilde{w} \in \mathbb{R}^d$ a context word embedding of dimension d

Motivation

Prob. and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$\frac{P(k\|ice)}{P(k\|steam)}$	8.9	8.5×10^{-2}	1.36	0.96

Motivation

Prob. and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$\frac{P(k\|ice)}{P(k\|steam)}$	8.9	8.5×10^{-2}	1.36	0.96

Motivation

Prob. and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$\frac{P(k\|ice)}{P(k\|steam)}$	8.9	8.5×10^{-2}	1.36	0.96

Motivation

Prob. and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$\frac{P(k\|ice)}{P(k\|steam)}$	8.9	8.5×10^{-2}	1.36	0.96

Motivation

Prob. and Ratio	$k = solid$	$k = gas$	$k = water$	$k = fashion$
$P(k\|ice)$	1.9×10^{-4}	6.6×10^{-5}	3.0×10^{-3}	1.7×10^{-5}
$P(k\|steam)$	2.2×10^{-5}	7.8×10^{-4}	2.2×10^{-3}	1.8×10^{-5}
$\frac{P(k\|ice)}{P(k\|steam)}$	8.9	8.5×10^{-2}	1.36	0.96

Derivation

$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

Derivation

$$F(w_i, w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

F should encode information in the ratio $\frac{P_{ik}}{P_{jk}}$.

Derivation

$$F(w_i - w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}}$$

Derivation

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

Derivation

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

Some more desiderata:

- F should be unchanged by exchanging $w \to \tilde{w}$ and $X \to X^T$

Derivation

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

Some more desiderata:

- F should be unchanged by exchanging $w \rightarrow \tilde{w}$ and $X \rightarrow X^T$

  This requires that

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F\left(w_i^T \tilde{w}_k\right)}{F\left(w_j^T \tilde{w}_k\right)}$$

$$\Rightarrow F(w_i^T \tilde{w}_k) = P_{ik}$$

Derivation

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

Some more desiderata:

- F should be unchanged by exchanging $w \to \tilde{w}$ and $X \to X^T$
  This requires that

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F\left(w_i^T \tilde{w}_k\right)}{F\left(w_j^T \tilde{w}_k\right)}$$

$$\Rightarrow F(w_i^T \tilde{w}_k) = P_{ik}$$

$$F\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{F\left(w_i^T \tilde{w}_k\right)}{F\left(w_j^T \tilde{w}_k\right)}$$

Derivation

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{P_{ik}}{P_{jk}}$$

Some more desiderata:

- F should be unchanged by exchanging $w \to \tilde{w}$ and $X \to X^T$
  This requires that

$$F\left((w_i - w_j)^T \tilde{w}_k\right) = \frac{F\left(w_i^T \tilde{w}_k\right)}{F\left(w_j^T \tilde{w}_k\right)}$$

$$\Rightarrow F(w_i^T \tilde{w}_k) = P_{ik}$$

$$\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)}$$

Derivation

$$\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)}$$

Derivation

$$\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)}$$

$$w_i^T \tilde{w}_k = \log P_{ik} = \log X_{ik} - \log X_i$$

Derivation

$$\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)}$$

$$w_i^T \tilde{w}_k = \log P_{ik} = \log X_{ik} - \log X_i$$

Derivation

$$\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) = \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)}$$

$$w_i^T \tilde{w}_k = \log P_{ik} = \log X_{ik} - \log X_i$$

$$w_i^T \tilde{w}_k = \log X_{ik} - b_i - \tilde{b}_k$$

Derivation

$$
\begin{aligned}
\exp\left(w_i^T \tilde{w}_k - w_j^T \tilde{w}_k\right) &= \frac{\exp\left(w_i^T \tilde{w}_k\right)}{\exp\left(w_j^T \tilde{w}_k\right)} \\
w_i^T \tilde{w}_k &= \log P_{ik} = \log X_{ik} - \log X_i \\
w_i^T \tilde{w}_k &= \log X_{ik} - b_i - \tilde{b}_k \\
w_i^T \tilde{w}_k + b_i + \tilde{b}_k &= \log X_{ik}
\end{aligned}
$$

Derivation

This suggests a least-squares objective function,

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

Derivation

This suggests a least-squares objective function, but...

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

Derivation

This suggests a least-squares objective function, but...

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

Derivation

This suggests a least-squares objective function, but...

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

Derivation

This suggests a least-squares objective function, but...

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

$$\Rightarrow J = \sum_{i,j=1}^{V} f\left(X_{ij}\right) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

where f has the following desiderata:

1. $f(0) = 0$
2. $f(x)$ should be non-decreasing so that rare co-occurrences are not overweighted.
3. $f(x)$ should be relatively small for large values of x, so that frequent co-occurrences are not overweighted.

Derivation

This suggests a least-squares objective function, but...

$$J = \sum_{i,j=1}^{V} \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

$$\Rightarrow J = \sum_{i,j=1}^{V} f\left(X_{ij}\right) \left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2$$

where $f(x) = \begin{cases} \left(\frac{x}{x_{\max}}\right)^{\alpha} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$

Weighting Function

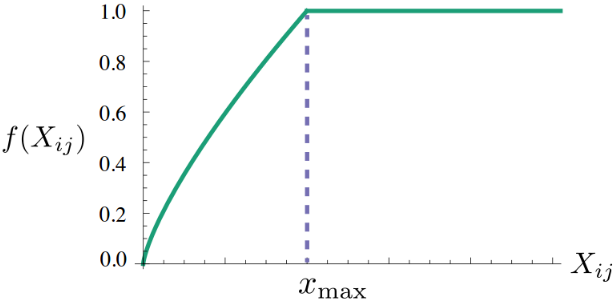


Figure 1: Weighting function f with $\alpha = 3/4$.

Optimization

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij}\right)^2$$

where $f(x) = \begin{cases} \left(\frac{x}{x_{\max}}\right)^{\alpha} & \text{if } x < x_{\max} \\ 1 & \text{otherwise} \end{cases}$

In this paper: $\alpha = \frac{3}{4}$ and $x_{\max} = 100$.

The model is trained using AdaGrad and stochastically sampling non-zero elements from X. An initial learning rate of .05 is used.

Word Analogies

a is to b as c to ?

$$w_b - w_a + w_c$$

Word Analogies

a is to b as c to ?
Paris is to *France* as *Tokyo* is to ?

$$w_b - w_a + w_c$$

Word Analogies

a is to b as c to ?
Paris is to *France* as *Tokyo* is to ?

$$\arg\max_{w'} \text{cosine-sim}(w_b - w_a + w_c, w')$$

Word Analogies – Results

Model	Dim.	Size	Sem.	Syn.	Tot.
ivLBL	100	1.5B	55.9	50.1	53.2
HPCA	100	1.6B	4.2	16.4	10.8
GloVe	100	1.6B	<u>67.5</u>	<u>54.3</u>	<u>60.3</u>
SG	300	1B	61	61	61
CBOW	300	1.6B	16.1	52.6	36.1
vLBL	300	1.5B	54.2	<u>64.8</u>	60.0
ivLBL	300	1.5B	65.2	63.0	64.0
GloVe	300	1.6B	<u>80.8</u>	61.5	<u>70.3</u>
SVD	300	6B	6.3	8.1	7.3
SVD-S	300	6B	36.7	46.6	42.1
SVD-L	300	6B	56.6	63.0	60.1
CBOW†	300	6B	63.6	<u>67.4</u>	65.7
SG†	300	6B	73.0	66.0	69.1
GloVe	300	6B	<u>77.4</u>	67.0	<u>71.7</u>
CBOW	1000	6B	57.3	68.9	63.7
SG	1000	6B	66.1	65.1	65.6
SVD-L	300	42B	38.4	58.2	49.2
GloVe	300	42B	**<u>81.9</u>**	**<u>69.3</u>**	**<u>75.0</u>**

Word Similarities

Humans scored similarity of word pairs.

word 1	word 2	human score (mean) (1-10)	cosine-similarity (-1, 1)
king	cabbage	0.23	0.11
king	queen	8.58	0.78
king	rook	5.92	0.25

Word Similarities

Humans scored similarity of word pairs.

word 1	word 2	human score (mean) (1-10)	cosine-similarity (-1, 1)
king	cabbage	0.23	0.11
king	queen	8.58	0.78
king	rook	5.92	0.25

Word Similarities

Humans scored similarity of word pairs.

word 1	word 2	human score (mean) (1-10)	cosine-similarity (-1, 1)
king	cabbage	0.23	0.11
king	queen	8.58	0.78
king	rook	5.92	0.25

Embeddings are evaluated by Spearman rank correlation of human scores to cosine similarity.

Word Similarities – Results

Model	Size	WS353	MC	RG	SCWS	RW
SVD	6B	35.3	35.1	42.5	38.3	25.6
SVD-S	6B	56.5	71.5	71.0	53.6	34.7
SVD-L	6B	65.7	<u>72.7</u>	75.1	56.5	37.0
CBOW†	6B	57.2	65.6	68.2	57.0	32.5
SG†	6B	62.8	65.2	69.7	<u>58.1</u>	37.2
GloVe	6B	<u>65.8</u>	<u>72.7</u>	<u>77.8</u>	53.9	<u>38.1</u>
SVD-L	42B	74.0	76.4	74.1	58.3	39.9
GloVe	42B	<u>**75.9**</u>	<u>**83.6**</u>	<u>**82.9**</u>	<u>**59.6**</u>	<u>**47.8**</u>
CBOW*	100B	68.4	79.6	75.4	59.4	45.5

Named Entity Recognition

NER is a sequence tagging task where the goal is to identify named entities:

Jim	bought	300	shares	of	Acme	Corp	.	in	2006	.
B-PER	O	O	O	O	B-ORG	I-ORG	I-ORG	O	B-TIME	O

Named Entity Recognition

NER is a sequence tagging task where the goal is to identify named entities:

Jim	bought	300	shares	of	Acme	Corp	.	in	2006	.
B-PER	O	O	O	O	B-ORG	I-ORG	I-ORG	O	B-TIME	O

Combined discrete features of existing system (Stanford NER).

Word embeddings were treated as additional features in a linear-chain CRF model.

Named Entity Recognition – Results

Model	Dev	Test	ACE	MUC7
Discrete	91.0	85.4	77.4	73.4
SVD	90.8	85.7	77.3	73.7
SVD-S	91.0	85.5	77.6	74.3
SVD-L	90.5	84.8	73.6	71.5
HPCA	92.6	**88.7**	81.7	80.7
HSMN	90.5	85.7	78.7	74.7
CW	92.2	87.4	81.7	80.2
CBOW	93.1	88.2	82.2	81.1
GloVe	**93.2**	88.3	**82.9**	**82.2**

The end! Thanks!