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
1 Introduction

2 Problem

3 GloVe Model

4 Experiments
Neural Language Models – Recurrent NNLM

\[ h_{t-1} \]

\[ w_t \]

\[ h_t \]

\[ o_{t+1} \]
Neural Language Models – Recurrent NNLM

\[ h_{t-1} \]

\[ w_t \]

\[ h_t \]

\[ o_{t+1} \]
Neural Language Models – Continuous BOW

\[
\begin{align*}
&w_{t-2} \\
&w_{t-1} \\
&w_{t+1} \\
&w_{t+2}
\end{align*}
\]

\[w_{\text{avg}}\]

\[o_t\]
Neural Language Models – Continuous BOW

\[ w_{t-2}, w_{t-1}, w_{t+1}, w_{t+2} \]

\[ w_{avg}, o_t \]
Linear Relationships

Semantic

\[ w_{king} - w_{man} + w_{woman} \approx w_{queen} \]

Syntactic

\[ w_{easy} - w_{easiest} + w_{luckiest} \approx w_{lucky} \]
Scalable Embedding Learning

Noise Contrastive Estimation

\[ w_{t-2} \]

\[ w_{t-1} \]

\[ w_{t+1} \]

\[ w_{avg} \]

\[ o_t \]

\[ w_{t+2} \]
Scalable Embedding Learning

Noise Contrastive Estimation – *no more normalization required!*

\[ w_{t-2} \]
\[ w_{t-1} \]
\[ w_{t+1} \]
\[ w_{avg} \]
\[ o_t \]

\[ w_{t+2} \]
Noise Contrastive Estimation – *no more normalization required!*
Lots of time spent scanning context windows to learn a distribution for \( P(w|the) \)
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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|\text{the}) \]

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

\[ P(w|th) \]

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

\[ P(w|\text{the}) \]

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.
1 Introduction

2 Problem

3 GloVe Model

4 Experiments
\[ J = \sum_{i,j=1}^{V} f(X_{ij}) \left( w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \]
- $X \in \mathbb{R}^{V \times V}$ 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
- $X_{i,j}$ 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 \( i \)
- $X \in \mathbb{R}^{V \times V}$ 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

<table>
<thead>
<tr>
<th>Prob. and Ratio</th>
<th>$k = \text{solid}$</th>
<th>$k = \text{gas}$</th>
<th>$k = \text{water}$</th>
<th>$k = \text{fashion}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k</td>
<td>\text{ice})$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$P(k</td>
<td>\text{steam})$</td>
<td>$2.2 \times 10^{-5}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\frac{P(k</td>
<td>\text{ice})}{P(k</td>
<td>\text{steam})}$</td>
<td>$8.9$</td>
<td>$8.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
## Motivation

<table>
<thead>
<tr>
<th>Prob. and Ratio</th>
<th>$k = solid$</th>
<th>$k = gas$</th>
<th>$k = water$</th>
<th>$k = fashion$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k</td>
<td>\text{ice})$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$P(k</td>
<td>\text{steam})$</td>
<td>$2.2 \times 10^{-5}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\frac{P(k</td>
<td>\text{ice})}{P(k</td>
<td>\text{steam})}$</td>
<td>$8.9$</td>
<td>$8.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
## Motivation

<table>
<thead>
<tr>
<th>Prob. and Ratio</th>
<th>$k = solid$</th>
<th>$k = gas$</th>
<th>$k = water$</th>
<th>$k = fashion$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k</td>
<td>ice)$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$P(k</td>
<td>steam)$</td>
<td>$2.2 \times 10^{-5}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\frac{P(k</td>
<td>ice)}{P(k</td>
<td>steam)}$</td>
<td>8.9</td>
<td>$8.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
Motivation

<table>
<thead>
<tr>
<th>Prob. and Ratio</th>
<th>$k = solid$</th>
<th>$k = gas$</th>
<th>$k = water$</th>
<th>$k = fashion$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k</td>
<td>\text{ice})$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$P(k</td>
<td>\text{steam})$</td>
<td>$2.2 \times 10^{-5}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\frac{P(k</td>
<td>\text{ice})}{P(k</td>
<td>\text{steam})}$</td>
<td>8.9</td>
<td>$8.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
## Motivation

<table>
<thead>
<tr>
<th>Prob. and Ratio</th>
<th>$k = solid$</th>
<th>$k = gas$</th>
<th>$k = water$</th>
<th>$k = fashion$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(k</td>
<td>ice)$</td>
<td>$1.9 \times 10^{-4}$</td>
<td>$6.6 \times 10^{-5}$</td>
<td>$3.0 \times 10^{-3}$</td>
</tr>
<tr>
<td>$P(k</td>
<td>steam)$</td>
<td>$2.2 \times 10^{-5}$</td>
<td>$7.8 \times 10^{-4}$</td>
<td>$2.2 \times 10^{-3}$</td>
</tr>
<tr>
<td>$\frac{P(k</td>
<td>ice)}{P(k</td>
<td>steam)}$</td>
<td>8.9</td>
<td>$8.5 \times 10^{-2}$</td>
</tr>
</tbody>
</table>
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

\[ F(w_i - w_j, \tilde{w}_k) = \frac{P_{ik}}{P_{jk}} \]
Some more desiderata:

\[ F((w_i - w_j)^T \tilde{w}_k) = \frac{P_{ik}}{P_{jk}} \]
Derivation

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

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

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

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

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

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

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

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

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

\[
w_i^T \tilde{w}_k = \log P_{ik} = \log X_{ik} - \log X_i
\]
\[ \exp (w_i^T \tilde{w}_k - w_j^T \tilde{w}_k) = \frac{\exp (w_i^T \tilde{w}_k)}{\exp (w_j^T \tilde{w}_k)} \]

\[ w_i^T \tilde{w}_k = \log P_{ik} = \log X_{ik} - \log X_i \]
\[
\exp \left( w_i^T \tilde{w}_k - w_j^T \tilde{w}_k \right) = \frac{\exp \left( w_i^T \tilde{w}_k \right)}{\exp \left( w_j^T \tilde{w}_k \right)}
\]

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

\[
w_i^T \tilde{w}_k = \log X_{ik} - b_i - \tilde{b}_k
\]
\[
\exp \left( w_i^T \tilde{w}_k - w_j^T \tilde{w}_k \right) = \frac{\exp \left( w_i^T \tilde{w}_k \right)}{\exp \left( w_j^T \tilde{w}_k \right)} \\
= \log P_{i,k} = \log X_{i,k} - \log X_i \\
= \log X_{i,k} - b_i - \tilde{b}_k \\
= \log X_{i,k} + b_i + \tilde{b}_k
\]
This suggests a least-squares objective function,

\[ J = \sum_{i,j=1}^{V} \left( w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \]
This suggests a least-squares objective function, but...

\[ J = \sum_{i,j=1}^{V} \left( w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \]
This suggests a least-squares objective function, but...

\[ J = \sum_{i,j=1}^{V} \left( w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \]
This suggests a least-squares objective function, but...

\[ J = \sum_{i,j=1}^{V} \left( w_i^T \tilde{w}_j + b_i + \tilde{b}_j - \log X_{ij} \right)^2 \]
This suggests a least-squares objective function, but...

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

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

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

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

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

where \( f(x) = \begin{cases} 
\left( \frac{x}{x_{\text{max}}} \right)^\alpha & \text{if } x < x_{\text{max}} \\
1 & \text{otherwise}
\end{cases} \)
Figure 1: Weighting function $f$ with $\alpha = 3/4$. 
Optimization

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

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

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

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

2 Problem

3 GloVe Model

4 Experiments
$a$ is to $b$ as $c$ to ?

$w_b - w_a + w_c$
Word Analogies

\[ a \text{ is to } b \text{ as } c \text{ to } ? \]

*Paris* is to *France* as *Tokyo* is to *?

\[ w_b - w_a + w_c \]
Word Analogies

\[ \begin{align*}
\text{a is to b as c to ?} \\
\text{Paris is to France as Tokyo is to ?} \\
\arg \max_{w'} \cosine-sim(w_b - w_a + w_c, w')
\end{align*} \]
## Word Analogies – Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ivLBL</td>
<td>100</td>
<td>1.5B</td>
<td>55.9</td>
<td>50.1</td>
<td>53.2</td>
</tr>
<tr>
<td>HPCA</td>
<td>100</td>
<td>1.6B</td>
<td>4.2</td>
<td>16.4</td>
<td>10.8</td>
</tr>
<tr>
<td>GloVe</td>
<td>100</td>
<td>1.6B</td>
<td>67.5</td>
<td>54.3</td>
<td>60.3</td>
</tr>
<tr>
<td>SG</td>
<td>300</td>
<td>1B</td>
<td>61</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>CBOW</td>
<td>300</td>
<td>1.6B</td>
<td>16.1</td>
<td>52.6</td>
<td>36.1</td>
</tr>
<tr>
<td>vLBL</td>
<td>300</td>
<td>1.5B</td>
<td>54.2</td>
<td>64.8</td>
<td>60.0</td>
</tr>
<tr>
<td>ivLBL</td>
<td>300</td>
<td>1.5B</td>
<td>65.2</td>
<td>63.0</td>
<td>64.0</td>
</tr>
<tr>
<td>GloVe</td>
<td>300</td>
<td>1.6B</td>
<td>80.8</td>
<td>61.5</td>
<td>70.3</td>
</tr>
<tr>
<td>SVD</td>
<td>300</td>
<td>6B</td>
<td>6.3</td>
<td>8.1</td>
<td>7.3</td>
</tr>
<tr>
<td>SVD-S</td>
<td>300</td>
<td>6B</td>
<td>36.7</td>
<td>46.6</td>
<td>42.1</td>
</tr>
<tr>
<td>SVD-L</td>
<td>300</td>
<td>6B</td>
<td>56.6</td>
<td>63.0</td>
<td>60.1</td>
</tr>
<tr>
<td>CBOW†</td>
<td>300</td>
<td>6B</td>
<td>63.6</td>
<td>67.4</td>
<td>65.7</td>
</tr>
<tr>
<td>SG†</td>
<td>300</td>
<td>6B</td>
<td>73.0</td>
<td>66.0</td>
<td>69.1</td>
</tr>
<tr>
<td>GloVe</td>
<td>300</td>
<td>6B</td>
<td>77.4</td>
<td>67.0</td>
<td>71.7</td>
</tr>
<tr>
<td>CBOW</td>
<td>1000</td>
<td>6B</td>
<td>57.3</td>
<td>68.9</td>
<td>63.7</td>
</tr>
<tr>
<td>SG</td>
<td>1000</td>
<td>6B</td>
<td>66.1</td>
<td>65.1</td>
<td>65.6</td>
</tr>
<tr>
<td>SVD-L</td>
<td>300</td>
<td>42B</td>
<td>38.4</td>
<td>58.2</td>
<td>49.2</td>
</tr>
<tr>
<td>GloVe</td>
<td>300</td>
<td>42B</td>
<td>81.9</td>
<td>69.3</td>
<td>75.0</td>
</tr>
</tbody>
</table>
Humans scored similarity of word pairs.

<table>
<thead>
<tr>
<th>word 1</th>
<th>word 2</th>
<th>human score (mean) (1-10)</th>
<th>cosine-similarity (-1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>king</td>
<td>cabbage</td>
<td>0.23</td>
<td>0.11</td>
</tr>
<tr>
<td>king</td>
<td>queen</td>
<td>8.58</td>
<td>0.78</td>
</tr>
<tr>
<td>king</td>
<td>rook</td>
<td>5.92</td>
<td>0.25</td>
</tr>
</tbody>
</table>
# Word Similarities

Humans scored similarity of word pairs.

<table>
<thead>
<tr>
<th>word 1</th>
<th>word 2</th>
<th>human score (mean) (1-10)</th>
<th>cosine-similarity (-1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>king</td>
<td>cabbage</td>
<td>0.23</td>
<td>0.11</td>
</tr>
<tr>
<td>king</td>
<td>queen</td>
<td>8.58</td>
<td>0.78</td>
</tr>
<tr>
<td>king</td>
<td>rook</td>
<td>5.92</td>
<td>0.25</td>
</tr>
</tbody>
</table>
Humans scored similarity of word pairs.

<table>
<thead>
<tr>
<th>word 1</th>
<th>word 2</th>
<th>human score (mean)</th>
<th>cosine-similarity (-1, 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>king</td>
<td>cabbage</td>
<td>0.23</td>
<td>0.11</td>
</tr>
<tr>
<td>king</td>
<td>queen</td>
<td>8.58</td>
<td>0.78</td>
</tr>
<tr>
<td>king</td>
<td>rook</td>
<td>5.92</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Embeddings are evaluated by Spearman rank correlation of human scores to cosine similarity.
## Word Similarities – Results

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Dev</th>
<th>Test</th>
<th>ACE</th>
<th>MUC7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete</td>
<td>91.0</td>
<td>85.4</td>
<td>77.4</td>
<td>73.4</td>
</tr>
<tr>
<td>SVD</td>
<td>90.8</td>
<td>85.7</td>
<td>77.3</td>
<td>73.7</td>
</tr>
<tr>
<td>SVD-S</td>
<td>91.0</td>
<td>85.5</td>
<td>77.6</td>
<td>74.3</td>
</tr>
<tr>
<td>SVD-L</td>
<td>90.5</td>
<td>84.8</td>
<td>73.6</td>
<td>71.5</td>
</tr>
<tr>
<td>HPCA</td>
<td>92.6</td>
<td><strong>88.7</strong></td>
<td>81.7</td>
<td>80.7</td>
</tr>
<tr>
<td>HSMN</td>
<td>90.5</td>
<td>85.7</td>
<td>78.7</td>
<td>74.7</td>
</tr>
<tr>
<td>CW</td>
<td>92.2</td>
<td>87.4</td>
<td>81.7</td>
<td>80.2</td>
</tr>
<tr>
<td>CBO%</td>
<td>93.1</td>
<td>88.2</td>
<td>82.2</td>
<td>81.1</td>
</tr>
<tr>
<td>GloVe</td>
<td><strong>93.2</strong></td>
<td><strong>88.3</strong></td>
<td><strong>82.9</strong></td>
<td><strong>82.2</strong></td>
</tr>
</tbody>
</table>
The end! Thanks!