These slides are distributed under the Creative Commons License.

DeepLearning.AI makes these slides available for educational purposes. You may not use or distribute these slides for commercial purposes. You may make copies of these slides and use or distribute them for educational purposes as long as you cite DeepLearning.AI as the source of the slides.

For the rest of the details of the license, see https://creativecommons.org/licenses/by-sa/2.0/legalcode
NLP and Word Embeddings

Word representation
Word representation

\[ V = [a, aaron, ..., zulu, <UNK>] \]

<table>
<thead>
<tr>
<th>Man</th>
<th>Woman</th>
<th>King</th>
<th>Queen</th>
<th>Apple</th>
<th>Orange</th>
</tr>
</thead>
<tbody>
<tr>
<td>5391</td>
<td>9853</td>
<td>4914</td>
<td>7157</td>
<td>456</td>
<td>6257</td>
</tr>
</tbody>
</table>

1-hot representation

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
\vdots \\
1 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
\vdots \\
1 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
1 \\
0 \\
\vdots \\
0 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
\vdots \\
0 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
\vdots \\
1 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

\[
\begin{bmatrix}
0 \\
0 \\
0 \\
0 \\
\vdots \\
0 \\
\vdots \\
0 \\
0 \\
0 \\
0 \\
0 \\
0
\end{bmatrix}
\]

I want a glass of orange [juice].

I want a glass of apple [_____].

Andrew Ng
### Featurized representation: word embedding

<table>
<thead>
<tr>
<th>Gender</th>
<th>Royal</th>
<th>Age</th>
<th>Food</th>
<th>Size</th>
<th>Cost</th>
<th>Smith</th>
<th>Verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.01</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
</tr>
<tr>
<td>Woman</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I want a glass of orange ______.</th>
</tr>
</thead>
<tbody>
<tr>
<td>I want a glass of apple ______.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Man (5391)</th>
<th>Woman (9853)</th>
<th>King (4914)</th>
<th>Queen (7157)</th>
<th>Apple (456)</th>
<th>Orange (6257)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.95</td>
<td>0.97</td>
<td>0.00</td>
<td>0.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.93</td>
<td>0.95</td>
<td>-0.01</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.69</td>
<td>0.03</td>
<td>-0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td>0.01</td>
<td>0.95</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Visualizing word embeddings

[van der Maaten and Hinton., 2008. Visualizing data using t-SNE]
NLP and Word Embeddings

Using word embeddings
Named entity recognition example

Sally Johnson is an orange farmer

Robert Lin is an apple farmer

→ 1B words - 100B words
→ 100K words

→ look words -> lookK words

a BRNN
Transfer learning and word embeddings

1. Learn word embeddings from large text corpus. (1-100B words)
   
   (Or download pre-trained embedding online.)

2. Transfer embedding to new task with smaller training set.
   (say, 100k words)

3. Optional: Continue to finetune the word embeddings with new data.
Relation to face encoding

\[ \mathbf{x}^{(i)} \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow f(x^{(i)}) \rightarrow \hat{y} \]

\[ \mathbf{x}^{(j)} \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow \rightarrow f(x^{(j)}) \]

| \[ \mathbf{v} | = 10,000 \]
| \[ Q_1, \ldots, Q_{10,000} \]

[Taigman et. al., 2014. DeepFace: Closing the gap to human level performance]
NLP and Word Embeddings

Properties of word embeddings
Analogies

<table>
<thead>
<tr>
<th></th>
<th>Man</th>
<th>Woman</th>
<th>King</th>
<th>Queen</th>
<th>Apple</th>
<th>Orange</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>-1</td>
<td>1</td>
<td>-0.95</td>
<td>0.97</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Royal</td>
<td>0.01</td>
<td>0.02</td>
<td>0.93</td>
<td>0.95</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Age</td>
<td>0.03</td>
<td>0.02</td>
<td>0.70</td>
<td>0.69</td>
<td>0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>Food</td>
<td>0.09</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.95</td>
<td>0.97</td>
</tr>
</tbody>
</table>

[Mikolov et. al., 2013, Linguistic regularities in continuous space word representations]
Analogies using word vectors

300D
Find word with arg max

eman − ewoman ≈ eking − ew

Sim(ew, eking − eman + ewoman)

30−75%
Cosine similarity

\[ \text{Sim}(u, v) = \frac{u^T v}{\|u\| \|v\|} \]

- Man:Woman as Boy:Girl
- Ottawa:Canada as Nairobi:Kenya
- Big:Bigger as Tall:Taller
- Yen:Japan as Ruble:Russia

Andrew Ng
NLP and Word Embeddings

Embedding matrix
Embedding matrix

In practice, use specialized function to look up an embedding.

Andrew Ng
Neural language model

I want a glass of orange

[Bengio et. al., 2003, A neural probabilistic language model]
Other context/target pairs

I want a glass of orange juice to go along with my cereal.

Context: Last 4 words.

4 words on left & right

Last 1 word

Nearby 1 word

a glass of orange ? - to go along with

Orange ?
glass ?

skip gram
NLP and Word Embeddings

Word2Vec
Skip-grams

I want a glass of orange juice to go along with my cereal.

[Mikolov et. al., 2013. Efficient estimation of word representations in vector space.]
Model

Vocab size = 10,000

Content $c$ ("orange") $\rightarrow$ Target $t$ ($\rightarrow$ "juice")

$O_c \rightarrow E \rightarrow e_c \rightarrow O \rightarrow \hat{y}$

$e_c = E o_c$

$O_T e_c$

Softmax. $p(t|c) = \frac{e^{O_T e_c}}{\sum_{j=1}^{10,000} e^{O_T e_c}}$

$O_T$ = parameter associated with word $t$

$\rightarrow \mathcal{L}(\hat{y}, y) = -\sum_{i=1}^{10,000} y_i \log \hat{y_i}$

Andrew Ng
Problems with softmax classification

\[
p(t | c) = \frac{e^{\theta_t^T c}}{\sum_{j=1}^{10,000} e^{\theta_j^T c}}
\]

How to sample the context \( c \)?

\( \Rightarrow \) the, of, a, and, to, ...

\( \Rightarrow \) orange, apple, durian

\( \Rightarrow \) P(c)
NLP and Word Embeddings

Negative sampling
Defining a new learning problem

I want a glass of orange juice to go along with my cereal.

[Mikolov et. al., 2013. Distributed representation of words and phrases and their compositionality]
Model

Softmax: \[ p(t|c) = \frac{e^{\theta^T_t e_c}}{\sum_{j=1}^{10,000} e^{\theta^T_j e_c}} \]

\[ P(y = 1 | c, t) = \sigma(\Theta_e^T e_c) \leq \]

context word target?

orange juice 1
orange king 0
orange book 0
the of 0

10,000 binary classification problem

Andrew Ng
### Selecting negative examples

<table>
<thead>
<tr>
<th>Context</th>
<th>Word</th>
<th>Target?</th>
</tr>
</thead>
<tbody>
<tr>
<td>orange</td>
<td>juice</td>
<td>1</td>
</tr>
<tr>
<td>orange</td>
<td>king</td>
<td>0</td>
</tr>
<tr>
<td>orange</td>
<td>book</td>
<td>0</td>
</tr>
<tr>
<td>orange</td>
<td>the</td>
<td>0</td>
</tr>
<tr>
<td>orange</td>
<td>of</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
P(\omega_i) = \frac{f(\omega_i)^{3/4}}{\sqrt[3]{\sum_{j=1}^{10,000} f(\omega_j)^{3/4}}} \quad \frac{1}{|V|} \]
NLP and Word Embeddings

GloVe word vectors
GloVe (global vectors for word representation)

I want a glass of orange juice to go along with my cereal.

\[ X_{ij} = \# \text{times } i \text{ appears in context of } j. \]

\[ X_{ji} = X_{ij} \leftarrow \]

[Pennington et. al., 2014. GloVe: Global vectors for word representation]
Model

Minimize

\[ \sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(x_{ij})(\Theta^T e_j + b_i + b_j - \log x_{ij})^2 \]

Weight term

\[ f(x_{ij}) = 0 \text{ or } x_{ij} = 0. \]

\[ \text{"O log 0" = 0} \]

This, is, \( a, a, \ldots \)

Diagonal

Therefore

\[ \Theta_i, e_j \text{ are symmetric} \]

\[ e_{\text{final}} = e_{\text{init}} + \Theta w \]

\[ e_{\text{final}} = \frac{e_{\text{init}} + \Theta w}{2} \]
A note on the featurization view of word embeddings

<table>
<thead>
<tr>
<th></th>
<th>Man</th>
<th>Woman</th>
<th>King</th>
<th>Queen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5391</td>
<td>9853</td>
<td>4914</td>
<td>7157</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Category</th>
<th>-1</th>
<th>1</th>
<th>-0.95</th>
<th>0.97</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>0.01</td>
<td>0.02</td>
<td>0.93</td>
<td>0.95</td>
</tr>
<tr>
<td>Royal</td>
<td>0.03</td>
<td>0.02</td>
<td>0.70</td>
<td>0.69</td>
</tr>
<tr>
<td>Age</td>
<td>0.09</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

minimize $\sum_{i=1}^{10,000} \sum_{j=1}^{10,000} f(X_{ij}) (\theta_i^T e_j + b_i - b_j' - \log X_{ij})^2$

Andrew Ng
NLP and Word Embeddings

Sentiment classification
The dessert is excellent. 🟢🟢🟢🟢☆

Service was quite slow. 🟢🟢🟢☆☆☆

Good for a quick meal, but nothing special. 🟢🟢🟢☆☆☆

Completely lacking in good taste, good service, and good ambience. 🟢☆☆☆☆☆

10,000 to 100,000 words
Simple sentiment classification model

The dessert is excellent

8928 2468 4694 3180

"Completely lacking in good taste, good service, and good ambience."
RNN for sentiment classification

\[ a^{<0>} \rightarrow a^{<1>} \rightarrow a^{<2>} \rightarrow a^{<3>} \rightarrow a^{<4>} \rightarrow \ldots \rightarrow a^{<10>} \]

\[ E \uparrow e_{1852} \quad e_{4966} \quad e_{4427} \quad e_{3882} \quad e_{330} \]

Completely lacking in good ambience.

softmax \[ \hat{y} \]

Andrew Ng
NLP and Word Embeddings

Debiasing word embeddings
The problem of bias in word embeddings

Man:Woman as King:Queen

Man:Computer_Programmer as Woman:Homemaker

Father:Doctor as Mother:Nurse

Word embeddings can reflect gender, ethnicity, age, sexual orientation, and other biases of the text used to train the model.

[Bolukbasi et. al., 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings]
Addressing bias in word embeddings

1. Identify bias direction.

2. Neutralize: For every word that is not definitional, project to get rid of bias.

3. Equalize pairs.

[Bolukbasi et. al., 2016. Man is to computer programmer as woman is to homemaker? Debiasing word embeddings]