Learning Effective and Interpretable Semantic Models using Non-Negative Sparse Embedding (NNSE)

Brian Murphy, Partha Talukdar, Tom Mitchell

Machine Learning Department
Carnegie Mellon University
Distributional Semantic Modeling

• Words are represented in a high dimensional vector space
Distributional Semantic Modeling

• Words are represented in a high dimensional vector space

• Long history:
  • (Deerwester et al., 1990), (Lin, 1998), (Turney, 2006), ...
Distributional Semantic Modeling

- Words are represented in a high dimensional vector space

- Long history:
  - (Deerwester et al., 1990), (Lin, 1998), (Turney, 2006), ...

- Although effective, these models are often not interpretable
Distributional Semantic Modeling

• Words are represented in a high dimensional vector space

• Long history:
  • (Deerwester et al., 1990), (Lin, 1998), (Turney, 2006), ...

• Although effective, these models are often not interpretable

<table>
<thead>
<tr>
<th>Model</th>
<th>Top 5 Words (per dimension)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD$_{300}$</td>
<td>well, long, if, year, watch</td>
</tr>
<tr>
<td></td>
<td>plan, engine, e, rock, very</td>
</tr>
<tr>
<td></td>
<td>get, no, features, music, via</td>
</tr>
<tr>
<td></td>
<td>features, by, links, free, down</td>
</tr>
<tr>
<td></td>
<td>works, sound, video, building, section</td>
</tr>
</tbody>
</table>

Examples of top 5 words from 5 randomly chosen dimensions from SVD$_{300}$
Why interpretable dimensions?
Semantic Decoding: (Mitchell et al., Science 2008)
Why interpretable dimensions?
Semantic Decoding: (Mitchell et al., Science 2008)

“motorbike”

Input
stimulus
word
Why interpretable dimensions?
Semantic Decoding: (Mitchell et al., Science 2008)

"motorbike" → (0.87, ride) → (0.29, see) → (0.00, rub) → (0.00, taste)

Input stimulus word

Semantic representation
Why interpretable dimensions?
Semantic Decoding: (Mitchell et al., Science 2008)

Input stimulus word

"motorbike"

Semantic representation

Mapping learned from fMRI data

(0.87, ride)
(0.29, see)
(0.00, rub)
(0.00, taste)
Why interpretable dimensions?
Semantic Decoding: (Mitchell et al., Science 2008)

Input stimulus word

"motorbike"

Semantic representation

Mapping learned from fMRI data

predicted activity for "motorbike"

(0.87, ride)

(0.29, see)

(0.00, rub)

(0.00, taste)
Why interpretable dimensions?  
(Mitchell et al., Science 2008)
Why interpretable dimensions? (Mitchell et al., Science 2008)

Fig. 4. Learned voxel activation signatures for 3 of the 25 semantic features, for participant P1 (top panels) and averaged over all nine participants (bottom panels). Just one horizontal z slice is shown for each. The semantic feature associated with the verb “eat” predicts substantial activity in right pars opercularis, which is believed to be part of the gustatory cortex. The semantic feature associated with “push” activates the right postcentral gyrus, which is believed to be associated with premotor planning. The semantic feature for the verb “run” activates the posterior portion of the right superior temporal sulcus, which is believed to be associated with the perception of biological motion.
Why interpretable dimensions? (Mitchell et al., Science 2008)

- Interpretable dimension reveals insightful brain activation patterns!
Why interpretable dimensions? 
(Mitchell et al., Science 2008)

- Interpretable dimension reveals insightful brain activation patterns!
- But, features in the semantic representation were based on 25 hand-selected verbs
Why interpretable dimensions? (Mitchell et al., Science 2008)

- Interpretable dimension reveals insightful brain activation patterns!
- But, features in the semantic representation were based on 25 hand-selected verbs
  - can’t represent arbitrary concepts
Why interpretable dimensions? (Mitchell et al., Science 2008)

- Interpretable dimension reveals insightful brain activation patterns!
- But, features in the semantic representation were based on 25 hand-selected verbs
  - can’t represent arbitrary concepts
  - need data-driven, broad coverage semantic representations
What is an Interpretable, Cognitively-plausible Representation?
What is an Interpretable, Cognitively-plausible Representation?

features

words
What is an Interpretable, Cognitively-plausible Representation?
What is an Interpretable, Cognitively-plausible Representation?

1. Compact representation: **Sparse**, many zeros
What is an Interpretable, Cognitively-plausible Representation?

1. Compact representation: **Sparse**, many zeros

2. Uneconomical to store negative (or inferable) characteristics: **Non-Negative**
What is an Interpretable, Cognitively-plausible Representation?

1. Compact representation: **Sparse**, many zeros

2. Uneconomical to store negative (or inferable) characteristics: **Non-Negative**

3. Meaningful Dimensions: **Coherent**
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th></th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand Tuned</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th>Hand Tuned</th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Corpus-derived (existing)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th></th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand Tuned</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corpus-derived (existing)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hand Coded</td>
<td></td>
<td></td>
<td></td>
<td>83.5</td>
<td></td>
</tr>
<tr>
<td>Corpus-derived</td>
<td></td>
<td></td>
<td></td>
<td>83.1</td>
<td></td>
</tr>
</tbody>
</table>

Prediction accuracy (on Neurosemantic Decoding)

(Murphy, Talukdar, Mitchell, StarSem 2012)
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th></th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand Tuned</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Corpus-derived (existing)</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th></th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hand Tuned</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Corpus-derived (existing)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NNSE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Our proposal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Properties of Different Semantic Representations

<table>
<thead>
<tr>
<th></th>
<th>Broad Coverage</th>
<th>Effective</th>
<th>Sparse</th>
<th>Non-Negative</th>
<th>Coherent</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hand Tuned</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Corpus-derived</strong> (existing)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>NNSE</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Our proposal
Non-Negative Sparse Embedding (NNSE)
Non-Negative Sparse Embedding (NNSE)

Input Matrix (Corpus Cooc + SVD)

X
Non-Negative Sparse Embedding (NNSE)

Input Matrix (Corpus Cooc + SVD)

$X$

$W_i$

input representation for word $w_i$
Non-Negative Sparse Embedding (NNSE)

Input Matrix (Corpus Cooc + SVD)

\[
X = A \times D
\]

input representation for word \( w_i \)
Non-Negative Sparse Embedding (NNSE)

$$\text{Input Matrix (Corpus Cooc + SVD)}$$

$$X$$

$$w_i$$

input representation for word $$w_i$$

$$= A \times D$$

$$\text{NNSE representation for word } w_i$$

$$A$$

$$D$$

$$\text{Basis}$$
Non-Negative Sparse Embedding (NNSE)

Input Matrix (Corpus Cooc + SVD)

\[ X = A \times D \]

- matrix A is non-negative
Non-Negative Sparse Embedding (NNSE)

Input Matrix (Corpus Cooc + SVD)

\[ X = A \times D \]

- matrix \( A \) is non-negative
- sparsity penalty on the rows of \( A \)
Non-Negative Sparse Embedding (NNSE)

- matrix $A$ is non-negative
- sparsity penalty on the rows of $A$
- alternating minimization between $A$ and $D$, using SPAMS package
NNSE Optimization

\[
X = A \times D
\]

\[
\underset{A \in \mathbb{R}^{m \times k}, D \in \mathbb{R}^{k \times n}}{\arg \min} \sum_{i=1}^{m} \left( \|X_{i,:} - A_{i,:} \times D \|^2 + \lambda \|A_{i,:}\|_1 \right)
\]

where,

\[
D_{i,:}D_{i,:}^T \leq 1, \ \forall 1 \leq i \leq k
\]

\[
A_{i,j} \geq 0, \ \forall 1 \leq i \leq m, \ \forall 1 \leq j \leq k
\]
Experiments
Experiments

• Three main questions
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
  2. What is the degree of sparsity of NNSE?
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
  2. What is the degree of sparsity of NNSE?
  3. Are NNSE dimensions coherent?
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
  2. What is the degree of sparsity of NNSE?
  3. Are NNSE dimensions coherent?

• Setup
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
  2. What is the degree of sparsity of NNSE?
  3. Are NNSE dimensions coherent?

• Setup
  • partial ClueWeb09, 16bn tokens, 540m sentences, 50m documents
Experiments

• Three main questions
  1. Are NNSE representations effective in practice?
  2. What is the degree of sparsity of NNSE?
  3. Are NNSE dimensions coherent?

• Setup
  • partial ClueWeb09, 16bn tokens, 540m sentences, 50m documents
  • dependency parsed using Malt parser
Baseline Representation: SVD
Baseline Representation: SVD

• For about 35k words (~adult vocabulary), extract
Baseline Representation: SVD

• For about 35k words (~adult vocabulary), extract
  • document co-occurrence
Baseline Representation: SVD

• For about 35k words (~adult vocabulary), extract
  • document co-occurrence
  • dependency features from the parsed corpus
Baseline Representation: SVD

- For about 35k words (~adult vocabulary), extract
  - document co-occurrence
  - dependency features from the parsed corpus
- Reduce dimensionality using SVD. Subsets of this reduced dimensional space is the baseline
Baseline Representation: SVD

• For about 35k words (~adult vocabulary), extract
  • document co-occurrence
  • dependency features from the parsed corpus

• Reduce dimensionality using SVD. Subsets of this reduced dimensional space is the baseline

• This is also the input (X) to NNSE
Baseline Representation: SVD

• For about 35k words (~adult vocabulary), extract
  • document co-occurrence
  • dependency features from the parsed corpus

• Reduce dimensionality using SVD. Subsets of this reduced dimensional space is the baseline

• This is also the input (X) to NNSE

• Other representations were also compared (e.g., LDA, Collobert and Weston, etc.), details in the paper
Is NNSE effective in Neurosemantic Decoding?
Is NNSE effective in Neurosemantic Decoding?
Is NNSE effective in Neurosemantic Decoding?

NNSE has similar peak performance as SVD
Does NNSE result in sparse semantic representation?
Does NNSE result in sparse semantic representation?

<table>
<thead>
<tr>
<th></th>
<th>SVD\textsubscript{300}</th>
<th>NNSE\textsubscript{50}</th>
<th>NNSE\textsubscript{300}</th>
<th>NNSE\textsubscript{1000}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparsity level (% of zeros)</td>
<td>0</td>
<td>81.94</td>
<td>90.39</td>
<td>99.95</td>
</tr>
<tr>
<td>Average number of words per dimension</td>
<td>35560.0</td>
<td>6422.4</td>
<td>3418.5</td>
<td>1818.2</td>
</tr>
<tr>
<td>Average number of dimensions per word</td>
<td>300.0</td>
<td>9.0</td>
<td>28.8</td>
<td>51.1</td>
</tr>
</tbody>
</table>
Does NNSE result in sparse semantic representation?

- NNSE is significantly sparser than SVD
Does NNSE result in sparse semantic representation?

<table>
<thead>
<tr>
<th></th>
<th>SVD_{300}</th>
<th>NNSE_{50}</th>
<th>NNSE_{300}</th>
<th>NNSE_{1000}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparsity level (% of zeros)</td>
<td>0</td>
<td>81.94</td>
<td>90.39</td>
<td>99.95</td>
</tr>
<tr>
<td>Average number of words per dimension</td>
<td>35560.0</td>
<td>6422.4</td>
<td>3418.5</td>
<td>1818.2</td>
</tr>
<tr>
<td>Average number of dimensions per word</td>
<td>300.0</td>
<td>9.0</td>
<td>28.8</td>
<td>51.1</td>
</tr>
</tbody>
</table>

- NNSE is significantly sparser than SVD
- Words per dimension is significantly lower in NNSE
Does NNSE result in sparse semantic representation?

- NNSE is significantly sparser than SVD
- Words per dimension is significantly lower in NNSE
- Growth in active dimensions per word is sub-linear in NNSE

<table>
<thead>
<tr>
<th></th>
<th>SVD_{300}</th>
<th>NNSE_{50}</th>
<th>NNSE_{300}</th>
<th>NNSE_{1000}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparsity level (% of zeros)</td>
<td>0</td>
<td>81.94</td>
<td>90.39</td>
<td>99.95</td>
</tr>
<tr>
<td>Average number of words per dimension</td>
<td>35560.0</td>
<td>6422.4</td>
<td>3418.5</td>
<td>1818.2</td>
</tr>
<tr>
<td>Average number of dimensions per word</td>
<td>300.0</td>
<td>9.0</td>
<td>28.8</td>
<td>51.1</td>
</tr>
</tbody>
</table>
Are NNSE Dimensions Coherent?
Are NNSE Dimensions Coherent?

- Word Intrusion Task  (Boyd-Graber et al., NIPS 2009)
Are NNSE Dimensions Coherent?

• Word Intrusion Task (Boyd-Graber et al., NIPS 2009)
  • for each dimension, select top ranked N (N=5) words
Are NNSE Dimensions Coherent?

• **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  • for each dimension, select top ranked N (N=5) words
  • intrude it with a low ranking word from this dimension
Are NNSE Dimensions Coherent?

• **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  • for each dimension, select top ranked $N$ ($N=5$) words
  • intrude it with a low ranking word from this dimension
  • intruding word should be high ranking in another dimension
Are NNSE Dimensions Coherent?

- **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  - for each dimension, select top ranked N (N=5) words
  - intrude it with a low ranking word from this dimension
  - intruding word should be high ranking in another dimension
  - ask a human to identify the intruding word
Are NNSE Dimensions Coherent?

• **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  • for each dimension, select top ranked N (N=5) words
  • intrude it with a low ranking word from this dimension
  • intruding word should be high ranking in another dimension
  • ask a human to identify the intruding word
  • repeat multiple times for each dimension, calculate precision
Are NNSE Dimensions Coherent?

- **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  - for each dimension, select top ranked $N$ ($N=5$) words
  - intrude it with a low ranking word from this dimension
  - intruding word should be high ranking in another dimension
  - ask a human to identify the intruding word
  - repeat multiple times for each dimension, calculate precision

- An intruded set from an NNSE$_{1000}$ dimension

  \{bathroom, closet, attic, balcony, quickly, toilet\}
Are NNSE Dimensions Coherent?

- **Word Intrusion Task** (Boyd-Graber et al., NIPS 2009)
  - for each dimension, select top ranked N (N=5) words
  - intrude it with a low ranking word from this dimension
  - intruding word should be high ranking in another dimension
  - ask a human to identify the intruding word
  - repeat multiple times for each dimension, calculate precision

- **An intruded set from an NNSE\textsubscript{1000} dimension**

  \{bathroom, closet, attic, balcony, quickly, toilet\}
Are NNSE Dimensions Coherent?
Are NNSE Dimensions Coherent?

The bar chart compares the precision of SVD and NNSE for different dimensions (k). The heights of the bars indicate the precision values:

- For k = 50, SVD precision is 44, NNSE precision is 72.
- For k = 300, SVD precision is 46.33, NNSE precision is 92.33.
- For k = 1000, SVD precision is 27.67, NNSE precision is 85.67.

The chart suggests that NNSE dimensions are generally more coherent than those of SVD, especially as the dimensionality increases.
Are NNSE Dimensions Coherent?

NNSE dimensions are significantly more coherent than SVD-based dimensions
### SVD and NNSE Dimensions: Examples

<table>
<thead>
<tr>
<th>Model</th>
<th>Top 5 Words (per dimension)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD(_{300})</td>
<td>well, long, if, year, watch plan, engine, e, rock, very get, no, features, music, via features, by, links, free, down works, sound, video, building, section</td>
</tr>
<tr>
<td>NNSE(_{1000})</td>
<td>inhibitor, inhibitors, antagonists, receptors, inhibition bristol, thames, southampton, brighton, poole delhi, india, bombay, chennai, madras pundits, forecasters, proponents, commentators, observers nosy, averse, leery, unsympathetic, snotty</td>
</tr>
</tbody>
</table>

Examples of top 5 words from 5 randomly chosen dimensions from SVD\(_{300}\) and NNSE\(_{1000}\)
### SVD and NNSE Dimensions: Examples

<table>
<thead>
<tr>
<th>Model</th>
<th>Top 5 Words (per dimension)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD$_{300}$</td>
<td>well, long, if, year, watch, plan, engine, e, rock, very, get, no, features, music, via, for</td>
</tr>
<tr>
<td>NNSE$_{1000}$</td>
<td>bristol, thames, southampton, brighton, poole, delhi, india, bombay, chennai, madras, pundits, forecasters, proponents, commentators, observers, nosy, averse, leery, unsympathetic, snotty</td>
</tr>
</tbody>
</table>

NNSE dimensions are significantly more coherent than SVD-based dimensions.

Examples of top 5 words from 5 randomly chosen dimensions from SVD$_{300}$ and NNSE$_{1000}$
Top 5 NNSE$_{1000}$ Dimensions for “apple”
## Top 5 NNSE\textsubscript{1000} Dimensions for “apple”

<table>
<thead>
<tr>
<th>Score</th>
<th>Dimensions</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>raspberry, peach, pear, mango, melon</td>
</tr>
<tr>
<td>0.26</td>
<td>ripper, aac, converter, vcd, rm</td>
</tr>
<tr>
<td>0.14</td>
<td>cpu, intel, mips, pentium, risc</td>
</tr>
<tr>
<td>0.13</td>
<td>motorola, lg, samsung, vodafone, alcatel</td>
</tr>
<tr>
<td>0.11</td>
<td>peaches, apricots, pears, cherries, blueberries</td>
</tr>
</tbody>
</table>
### Top 5 NNSE\(^{1000}\) Dimensions for “apple”

<table>
<thead>
<tr>
<th></th>
<th>Content</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>raspberry, peach, pear, mango, melon</td>
<td>Fruit</td>
</tr>
<tr>
<td>0.26</td>
<td>ripper, aac, converter, vcd, rm</td>
<td>Music</td>
</tr>
<tr>
<td>0.14</td>
<td>cpu, intel, mips, pentium, risc</td>
<td>Processor</td>
</tr>
<tr>
<td>0.13</td>
<td>motorola, lg, samsung, vodafone, alcatel</td>
<td>Company</td>
</tr>
<tr>
<td>0.11</td>
<td>peaches, apricots, pears, cherries, blueberries</td>
<td>Fruit</td>
</tr>
</tbody>
</table>
Top 5 NNSE\textsubscript{1000} Dimensions for “apple”

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.40</td>
<td>raspberry, peach, pear, mango, melon</td>
</tr>
<tr>
<td>0.26</td>
<td>ripper, aac, converter, vcd, rm</td>
</tr>
<tr>
<td>0.14</td>
<td>cpu, intel, mips, pentium, risc</td>
</tr>
<tr>
<td>0.13</td>
<td>motorola, lg, samsung, vodafone, alcatel</td>
</tr>
<tr>
<td>0.11</td>
<td>peaches, apricots, pears, cherries, blueberries</td>
</tr>
</tbody>
</table>

Different senses of the word are not mixed, each dimension corresponds to only one sense of “apple”!
Conclusion

• Non-Negative Sparse Embedding (NNSE)
  • broad coverage, sparse, non-negative
  • interpretable, effective in practice
  • probably first semantic model with all these desirable traits
Conclusion

• Non-Negative Sparse Embedding (NNSE)
  • broad coverage, sparse, non-negative
  • interpretable, effective in practice
  • probably first semantic model with all these desirable traits

• Exploited large text corpus, including deep linguistic features (e.g., dependency parses)
Conclusion

• Non-Negative Sparse Embedding (NNSE)
  • broad coverage, sparse, non-negative
  • interpretable, effective in practice
  • probably first semantic model with all these desirable traits

• Exploited large text corpus, including deep linguistic features (e.g., dependency parses)

• Future work
  • multi-word extension; using NNSE representations in non-neurosemantic domains (e.g., NELL)
Acknowledgment

• Khaled El-Arini (CMU) and Min Xu (CMU)
• Sheshadri Sridharan (CMU), Marco Baroni (Trento)
• Justin Betteridge (CMU), Jamie Callan (CMU)
• DARPA, Google, NIH
Thank You!

NNSE embeddings available at:
http://www.cs.cmu.edu/~bmurphy/NNSE/