Measuring Semantic Distance
using Distributional Profiles of Concepts

Saif Mohammad
Department of Computer Science
University of Toronto

Grateful acknowledgments: Graeme Hirst (advisor and co-author); Iryna Gurevych, Torsten Zesch, and Philip Resnik (co-authors); Rada Mihalcea, Renee Miller, Gerald Penn, Suzanne Stevenson, University of Toronto (especially the CL group), and NSERC.
Semantic Distance

SALSA  DANCE

CLOWN  BRIDGE

A measure of how close or distant two units of language are in terms of their meaning
Why measure semantic distance?

- Natural language processing is teeming with semantic-distance problems:
  - Machine translation

*You know a person by the company they keep*

*Das Wesen eines Menschen erkennt man an der Gesellschaft, mit der er sich umgibt*
Why measure semantic distance?

- Natural language processing is teeming with semantic-distance problems:
  - Word sense disambiguation

_Hermione cast a bewitching spell_

CHARM OR INCANTATION
Why measure semantic distance?

- Natural language processing is teeming with semantic-distance problems:
  - Speech recognition, real-word spelling correction

...interest...money...\textit{band}...loan...

\textit{bank} or \textit{bond}
Knowledge source–based semantic measures

- Structure of a network or resource
  - The nodes represent senses or concepts

- Drawbacks
  - Resource bottleneck
  - Not easily domain-adaptable
  - Accuracy on pairs other than noun–noun is poor
  - Relatedness estimation is poor
Corpus-based distributional measures

- Words in similar contexts are close.
  - Distributional profile (DP) of a word: strength of association of the word with co-occurring words in text
DP of a word

DP of fusion

heat 0.16
hydrogen 0.16
energy 0.13
hot 0.09
light 0.09
space 0.04
gravity 0.03
pressure 0.03
## DPs of words

<table>
<thead>
<tr>
<th>DP of <em>star</em></th>
<th>DP of <em>fusion</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>space</em> 0.21</td>
<td><em>heat</em> 0.16</td>
</tr>
<tr>
<td><em>movie</em> 0.16</td>
<td><em>hydrogen</em> 0.16</td>
</tr>
<tr>
<td><em>famous</em> 0.15</td>
<td><em>energy</em> 0.13</td>
</tr>
<tr>
<td><em>light</em> 0.12</td>
<td><em>hot</em> 0.09</td>
</tr>
<tr>
<td><em>rich</em> 0.11</td>
<td><em>light</em> 0.09</td>
</tr>
<tr>
<td><em>heat</em> 0.08</td>
<td><em>space</em> 0.04</td>
</tr>
<tr>
<td><em>planet</em> 0.07</td>
<td><em>gravity</em> 0.03</td>
</tr>
<tr>
<td><em>hydrogen</em> 0.07</td>
<td><em>pressure</em> 0.03</td>
</tr>
</tbody>
</table>
Distance between two words

**DP of star**
- space 0.21
- movie 0.16
- famous 0.15
- light 0.12
- rich 0.11
- heat 0.08
- planet 0.07
- hydrogen 0.07

**DP of fusion**
- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
- gravity 0.03
- pressure 0.03
Distance between two words

<table>
<thead>
<tr>
<th>DP of <em>star</em></th>
<th>DP of <em>fusion</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>space 0.21</td>
<td>heat 0.16</td>
</tr>
<tr>
<td>movie 0.16</td>
<td>hydrogen 0.16</td>
</tr>
<tr>
<td>famous 0.15</td>
<td>energy 0.13</td>
</tr>
<tr>
<td>light 0.12</td>
<td>hot 0.09</td>
</tr>
<tr>
<td>rich 0.11</td>
<td>light 0.09</td>
</tr>
<tr>
<td>heat 0.08</td>
<td>space 0.04</td>
</tr>
<tr>
<td>planet 0.07</td>
<td>gravity 0.03</td>
</tr>
<tr>
<td>hydrogen 0.07</td>
<td>pressure 0.03</td>
</tr>
</tbody>
</table>
Distance between two words

DP of *star*

- space 0.21
- movie 0.16
- famous 0.15
- light 0.12
- rich 0.11
- heat 0.08
- planet 0.07
- hydrogen 0.07

DP of *fusion*

- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
- gravity 0.03
- pressure 0.03
Distance between two words

DP of **star**

- space 0.21
- movie 0.16
- famous 0.15
- light 0.12
- rich 0.11
- heat 0.08
- planet 0.07
- hydrogen 0.07

DP of **fusion**

- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
- gravity 0.03
- pressure 0.03
Distance between two words

DP of \textit{star}

- space 0.21
- movie 0.16
- famous 0.15
- light 0.12
- rich 0.11
- heat 0.08
- planet 0.07
- hydrogen 0.07

DP of \textit{fusion}

- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
- gravity 0.03
- pressure 0.03
Distributional measures of word-distance

- Words in similar contexts are close.
  - Distributional profile (DP) of a word: strength of association of the word with co-occurring words (text)
  - Distributional measure: distance between DPs
    - Cosine, Lin, α-skew divergence

- Drawback
  - Poor accuracy (albeit higher coverage)
    - Conflation of word senses
Problem with distributional word-distance measures

<table>
<thead>
<tr>
<th>DP of <em>star</em></th>
<th>DP of <em>fusion</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>space</em> 0.21</td>
<td><em>heat</em> 0.16</td>
</tr>
<tr>
<td><em>movie</em> 0.16</td>
<td><em>hydrogen</em> 0.16</td>
</tr>
<tr>
<td><em>famous</em> 0.15</td>
<td><em>energy</em> 0.13</td>
</tr>
<tr>
<td><em>light</em> 0.12</td>
<td><em>hot</em> 0.09</td>
</tr>
<tr>
<td><em>rich</em> 0.11</td>
<td><em>light</em> 0.09</td>
</tr>
<tr>
<td><em>heat</em> 0.08</td>
<td><em>space</em> 0.04</td>
</tr>
<tr>
<td><em>planet</em> 0.07</td>
<td><em>gravity</em> 0.03</td>
</tr>
<tr>
<td><em>hydrogen</em> 0.07</td>
<td><em>pressure</em> 0.03</td>
</tr>
</tbody>
</table>
Problem with distributional word-distance measures

DP of star

- space 0.21
- movie 0.16
- famous 0.15
- light 0.12
- rich 0.11
- heat 0.08
- planet 0.07
- hydrogen 0.07

DP of fusion

- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
- gravity 0.03
- pressure 0.03

Word sense ambiguity reduces accuracy of distance measures
Shared limitations

- Precomputing all distances is computationally expensive
  - WordNet-based measures:
    - $117,000 \times 117,000$ sense–sense distance matrix
  - Distributional measures:
    - $100,000 \times 100,000$ word–word distance matrix

- Monolingual
A new hybrid approach

- Combines a knowledge source with text
  - Thesaurus categories: concepts/coarse senses
  - Most published thesauri: around 1000 categories

- Profiles concepts (rather than words)
  - Uses sets of words to represent each concept
  - Creates profiles using bootstrapping
Features

- Can be used in real-time applications
  - Concept–concept distance matrix: only $1000 \times 1000$
- Accurate for all pos–pos pairs
  - Not just noun–noun
- Capable of giving both similarity and relatedness values
- Easily domain adaptable
- Cross-lingual
Problem with distributional word-distance measures

DP of *star*

- *space* 0.21
- *movie* 0.16
- *famous* 0.15
- *light* 0.12
- *rich* 0.11
- *heat* 0.08
- *planet* 0.07
- *hydrogen* 0.07

Word sense ambiguity reduces accuracy of distance measures
Solution: tease out the senses

**star**

- space
- movie
- famous
- light
- rich
- heat
- planet
- hydrogen
Solution: tease out the senses

star

space

light

heat

planet

hydrogen

Profile the senses separately.
Distributional profiles of concepts

DPs of the concepts referred to by *star*:

<table>
<thead>
<tr>
<th>DP of CELESTIAL BODY</th>
<th>DP of CELEBRITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>space 0.36</td>
<td>famous 0.24</td>
</tr>
<tr>
<td>light 0.27</td>
<td>movie 0.14</td>
</tr>
<tr>
<td>heat 0.11</td>
<td>rich 0.14</td>
</tr>
<tr>
<td>planet 0.07</td>
<td>fan 0.10</td>
</tr>
<tr>
<td>hydrogen 0.06</td>
<td>hot 0.04</td>
</tr>
<tr>
<td>hot 0.01</td>
<td>fashion 0.01</td>
</tr>
</tbody>
</table>
Distributional profiles of concepts

DPs of the concepts referred to by *star*:

**DP of CELESTIAL BODY**
*(celestial body, star, sun,...)*
- space 0.36
- light 0.27
- heat 0.11
- planet 0.07
- hydrogen 0.06
- hot 0.01

**DP of CELEBRITY**
*(celebrity, hero, star,...)*
- famous 0.24
- movie 0.14
- rich 0.14
- fan 0.10
- hot 0.04
- fashion 0.01
Distance: *star and fusion*

DP of **FUSION**

*(atomic reaction, fusion, thermonuclear reaction,...)*

- heat 0.16
- hydrogen 0.16
- energy 0.13
- hot 0.09
- light 0.09
- space 0.04
## Distance: \textit{star} and \textit{fusion}

<table>
<thead>
<tr>
<th>DP of \textbf{CELEBRITY}</th>
<th>DP of \textbf{FUSION}</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{(celebrity, hero, star,...)}</td>
<td>\textit{(atomic reaction, fusion, thermonuclear reaction,...)}</td>
</tr>
<tr>
<td>famous 0.24</td>
<td>heat 0.16</td>
</tr>
<tr>
<td>movie 0.14</td>
<td>hydrogen 0.16</td>
</tr>
<tr>
<td>rich 0.14</td>
<td>energy 0.13</td>
</tr>
<tr>
<td>fan 0.10</td>
<td>hot 0.09</td>
</tr>
<tr>
<td>hot 0.04</td>
<td>light 0.09</td>
</tr>
<tr>
<td>fashion 0.01</td>
<td>space 0.04</td>
</tr>
</tbody>
</table>

First, consider the \textbf{CELEBRITY} sense of \textit{star}.
## Distance: *star* and *fusion*

<table>
<thead>
<tr>
<th>DP of <strong>CELEBRITY</strong></th>
<th>DP of <strong>FUSION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(celebrity, hero, star,...)</em></td>
<td><em>(atomic reaction, fusion, thermonuclear reaction,...)</em></td>
</tr>
<tr>
<td>famous 0.24</td>
<td>heat 0.16</td>
</tr>
<tr>
<td>movie 0.14</td>
<td>hydrogen 0.16</td>
</tr>
<tr>
<td>rich 0.14</td>
<td>energy 0.13</td>
</tr>
<tr>
<td>fan 0.10</td>
<td>hot 0.09</td>
</tr>
<tr>
<td><strong>hot</strong> 0.04</td>
<td>light 0.09</td>
</tr>
<tr>
<td>fashion 0.01</td>
<td>space 0.04</td>
</tr>
</tbody>
</table>

First, consider the **CELEBRITY** sense of *star*.

- Distributionally **NOT** close
Distance: *star* and *fusion*

**DP of FUSION**

*(atomic reaction, fusion, thermonuclear reaction,...)*

heat 0.16
hydrogen 0.16
energy 0.13
hot 0.09
light 0.09
space 0.04
Distance: *star* and *fusion*

<table>
<thead>
<tr>
<th><strong>DP of CELESTIAL BODY</strong></th>
<th><strong>DP of FUSION</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(celestial body, star, sun...)</em></td>
<td><em>(atomic reaction, fusion, thermonuclear reaction,...)</em></td>
</tr>
<tr>
<td>space 0.36</td>
<td>heat 0.16</td>
</tr>
<tr>
<td>light 0.27</td>
<td>hydrogen 0.16</td>
</tr>
<tr>
<td>heat 0.11</td>
<td>energy 0.13</td>
</tr>
<tr>
<td>planet 0.07</td>
<td>hot 0.09</td>
</tr>
<tr>
<td>hydrogen 0.07</td>
<td>light 0.09</td>
</tr>
<tr>
<td>hot 0.07</td>
<td>space 0.04</td>
</tr>
</tbody>
</table>

Then, consider the **CELESTIAL BODY** sense of *star*. 
Distance: *star* and *fusion*

<table>
<thead>
<tr>
<th>DP of CELESTIAL BODY</th>
<th>DP of FUSION</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(celestial body, star, sun...)</em></td>
<td><em>(atomic reaction, fusion, thermonuclear reaction,...)</em></td>
</tr>
<tr>
<td>space 0.36</td>
<td>heat 0.16</td>
</tr>
<tr>
<td>light 0.27</td>
<td>hydrogen 0.16</td>
</tr>
<tr>
<td>heat 0.11</td>
<td>energy 0.13</td>
</tr>
<tr>
<td>planet 0.07</td>
<td>hot 0.09</td>
</tr>
<tr>
<td>hydrogen 0.07</td>
<td>light 0.09</td>
</tr>
<tr>
<td>hot 0.07</td>
<td>space 0.04</td>
</tr>
</tbody>
</table>

Then, consider the CELESTIAL BODY sense of *star*.

- Distributionally **close**
- Word sense ambiguity **NOT** a problem
Ranking word pairs
(Monolingual)

![Bar chart showing correlation between word-distance and concept-distance for different distributional measures: ASD, Cos, JSN, and Lin.](image-url)
Correcting spelling errors
(Monolingual)

![Bar chart showing correction ratios for different distributional measures: ASD, Cos, JSN, Lin. Orange bars represent word-distance, and purple bars represent concept-distance.](Image)
But... Application of distance algorithms in most languages is hindered by a lack of high-quality linguistic resources.
So: Make it cross-lingual

- Determining distance in a resource-poor language
  - Combine its text with a thesaurus from a (possibly resource-rich) language
  - Largely alleviates the knowledge-source bottleneck
  - Use a bilingual lexicon
  - Without parallel corpora or sense-annotated data

- Experiments: German as a “resource-poor” language
Cross-lingual links

German words \( w^{de} \)

\[ \text{Stern} \quad \rightarrow_{w^{de}} \quad \text{Bank} \]
Cross-lingual links

German words $w^{de}$

English translations $w^{en}$ (German–English lexicon)
Cross-lingual links

German words $w^{de}$
English translations $w^{en}$ (German–English lexicon)
English concepts $c^{en}$ (English thesaurus)
The concepts of CELEBRITY, RIVER BANK and JUDICIARY are semantically unrelated to Stern and Bank.
Losing the English words

Measuring Semantic Distance using Distributional Profiles of Concepts. Saif Mohammad.
Losing the English words

\[ \text{CELEBRITY} \quad \text{CELESTIAL} \quad \text{RIVER} \quad \text{FINANCIAL} \quad \text{JUDICIARY} \]

\[ \begin{align*}
\text{BODY} & \quad \text{BANK} & \quad \text{INSTITUTION} & \quad \text{FURNITURE} \\
Stern & \quad & \text{Bank} & \quad \end{align*} \]

\[ C^\text{en} \]

\[ w^\text{de} \]
Losing the English words

Cross-lingual candidate senses of German words
*Stern* and *Bank*
Cross-lingual DPCs

Cross-lingual DPs of the concepts referred to by *star*:

**DP of CELESTIAL BODY**
(celestial body, star, sun,...)

**DP of CELEBRITY**
(celebrity, hero, star,...)
Cross-lingual DPCs

Cross-lingual DPs of the concepts referred to by *star*:

<table>
<thead>
<tr>
<th>DP of CELESTIAL BODY (celestial body, star, sun,...)</th>
<th>DP of CELEBRITY (celebrity, hero, star,...)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raum 0.36</td>
<td>berühmt 0.24</td>
</tr>
<tr>
<td>Licht 0.27</td>
<td>Film 0.14</td>
</tr>
<tr>
<td>Hitze 0.11</td>
<td>reich 0.14</td>
</tr>
<tr>
<td>Planet 0.07</td>
<td>Fan 0.10</td>
</tr>
<tr>
<td>Wasserstoff 0.06</td>
<td>heiß 0.04</td>
</tr>
<tr>
<td>heiß 0.01</td>
<td>Mode 0.01</td>
</tr>
</tbody>
</table>
Ranking word pairs (Cross-lingual)

![Bar chart showing correlation between monolingual and cross-lingual measures for different datasets and measures: Spearman's and Pearson's correlations for Gur65 and Gur350 datasets.](image)

Measuring Semantic Distance using Distributional Profiles of Concepts. Saif Mohammad. 45
Solving word choice problems
(Cross-lingual)

![Bar chart showing comparison between monolingual and cross-lingual precision, recall, and F-measure.](chart.png)
Distance between a concept and its context

word ... word ... target word ... word ... word
Distance between a concept and its context

CONCEPT2

CONCEPT1

word ... word ... target word ... word ... word
Distance between a concept and its context

CONCEPT2

CONCEPT1

word ... word ... target word ... word ... word

Measuring Semantic Distance using Distributional Profiles of Concepts. Saif Mohammad. 49
Distance between a concept and its context

CONCEPT2

CONCEPT1

word ... word ... target word ... word ... word
Distance between a concept and its context

Word sense dominance and word sense disambiguation:

- Obviate the need of sense-annotated data
Word sense dominance

Mean distance below upper bound = 0.02
Unsupervised Naïve Bayes word sense classifier

- Estimated probabilities from the DPC
- Took part in SemEval-07’s:
  - English Lexical Sample Task
    - Only one percentage point behind the best unsupervised system
  - Multilingual Chinese–English Lexical Sample Task
    - Placed clear first among unsupervised systems
Accomplishments (1)

- Performed a qualitative and quantitative comparison of WordNet-based and distributional measures

- Identified significant limitations of state-of-the-art approaches to measuring semantic distance
  - Word sense ambiguity
  - A hurdle for distributional measures
Accomplishments (2)

- Proposed a new hybrid approach to semantic distance
  - Combines text with a thesaurus
  - Models concepts (rather than words)
  - Uses thesaurus categories as very coarse senses
Accomplishments (3)

- Extensive evaluation
  - Monolingual
    - By combining English text with an English thesaurus
      - Ranked word pairs
      - Corrected real-word spelling errors
      - Determined word sense dominance
      - Did word sense disambiguation
Accomplishments (4)

- Extensive evaluation (continued)
  - Cross-lingual
    - By combining German text with an English thesaurus
      - Ranked word pairs and solving word-choice problems in German
    - By combining Chinese text with an English thesaurus
      - Identified the English translations of Chinese words from their contexts
Future work

- Adding cross-lingual semantic distance as a feature to a state-of-the-art MT system (with Philip Resnik)
- Cross-lingual document clustering
- Cross-lingual information retrieval
- Cross-lingual summarization (with Bonnie Dorr)
- Determining paraphrases, lexical entailment, and contradictions (with Bonnie Dorr)
- Determining cognates using semantic distance between words in different languages (with Greg Kondrak)
- Porting the approach to Wikipedia (with Torsten Zesch and Iryna Gurevych)
Conclusions (1)

- Distributional profiles of concepts can be used to infer their semantic properties, and indeed estimate semantic distance.
- Cross-lingual DPCs allow for a seamless transition from words in one language to concepts in another.
Conclusions (2)

- Distributional measures of concept-distance are markedly superior to previous approaches.
Conclusions (2)

- Distributional measures of concept-distance are markedly superior to previous approaches.
  - Works well for all pos pairs
Conclusions (2)

- Distributional measures of concept-distance are markedly superior to previous approaches.
  - Works well for all pos pairs
  - Gives both relatedness and similarity
Conclusions (2)

- Distributional measures of concept-distance are markedly superior to previous approaches.
  - Works well for all pos pairs
  - Gives both relatedness and similarity
  - Domain adaptable
Conclusions (2)

- **Distributional measures of concept-distance** are markedly superior to previous approaches.
  - Works well for all pos pairs
  - Gives both relatedness and similarity
  - Domain adaptable
  - Can be used in real-time systems
Conclusions (2)

- **Distributional measures of concept-distance** are markedly superior to previous approaches.
  - Works well for all pos pairs
  - Gives both relatedness and similarity
  - Domain adaptable
  - Can be used in real-time systems
- **Cross-lingual**
  - Solve problems in a one language using a knowledge source from another
  - Solve problems that involve multiple languages