Word Embeddings and Gender Sterotypes in Swedish and English

Rasmus Précenth

February 13, 2019

Word Embeddings

Words as vectors for use in machine learning, AI and natural language processing.

$$V \quad \rightarrow \quad \mathbb{R}^d$$

Word Embeddings

Words as vectors for use in machine learning, AI and natural language processing.

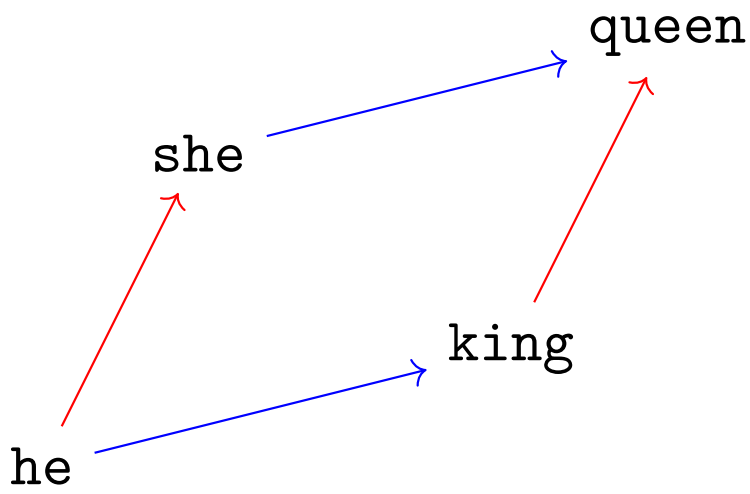
V	$\to$	$\mathbb{R}^d$
`he`	$\mapsto$	$(-0.22, 0.21, -0.04)$
`she`	$\mapsto$	$(-0.37, -0.16, 0.05)$
`is`	$\mapsto$	$(-0.56, 0.00, -0.07)$
`the`	$\mapsto$	$(-0.61, -0.03, -0.05)$
`and`	$\mapsto$	$(-0.30, 0.02, 0.20)$
`queen`	$\mapsto$	$(-0.21, -0.20, -0.05)$
`king`	$\mapsto$	$(-0.37, 0.18, 0.02)$

Word Embeddings and Sexism

He is to *king* as *she* is to *queen*.

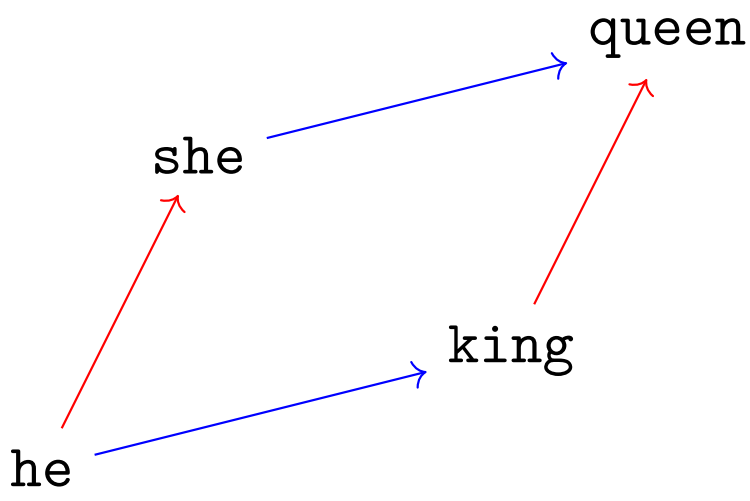
Word Embeddings and Sexism

He is to *king* as *she* is to *queen*.

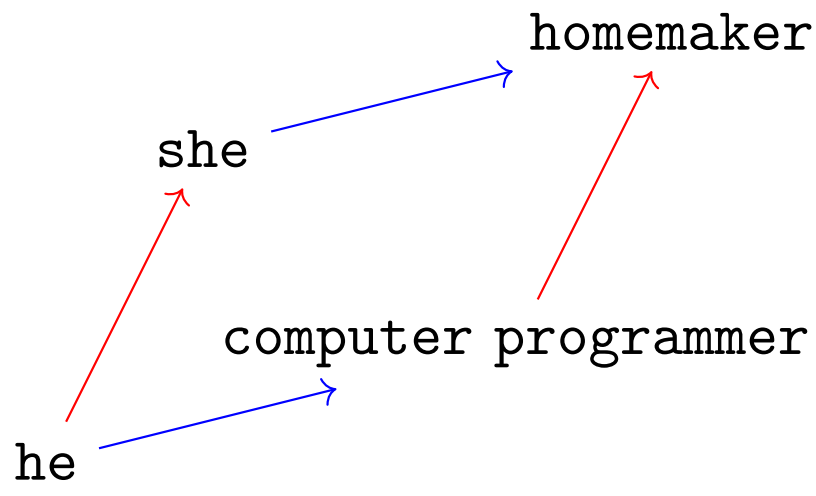


Word Embeddings and Sexism

He is to *king* as *she* is to *queen*.

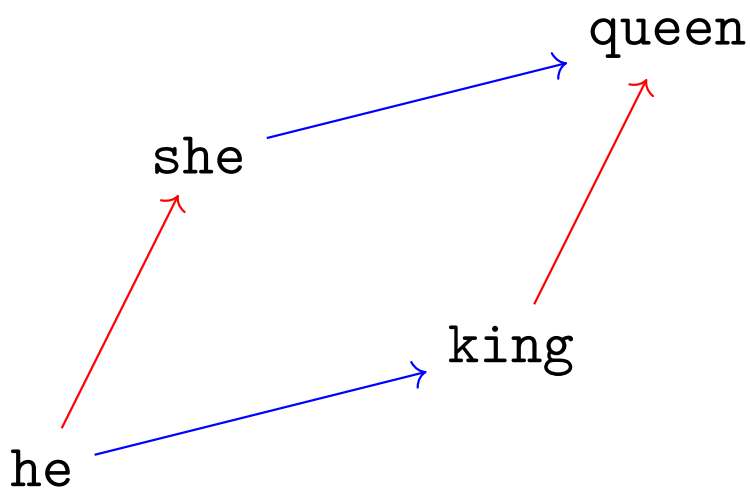


He is to *computer programmer* as *she* is to *homemaker*.

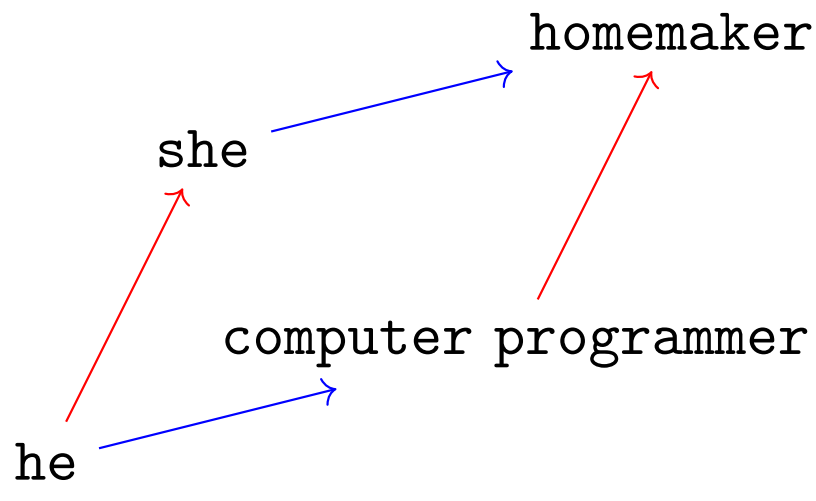


Word Embeddings and Sexism

He is to *king* as *she* is to *queen*.



He is to *computer programmer* as *she* is to *homemaker*.



Whoops!

Outline

Outline

Constructing Word Embeddings

Distributional Hypothesis (Harris 1954)

Words are determined by the company they keep.

Constructing Word Embeddings

Distributional Hypothesis (Harris 1954)

Words are determined by the company they keep.

Translation

- Lots and lots of data.

Constructing Word Embeddings

Distributional Hypothesis (Harris 1954)

Words are determined by the company they keep.

Translation

- Lots and lots of data.
- Count co-occurances.

Constructing Word Embeddings

Distributional Hypothesis (Harris 1954)

Words are determined by the company they keep.

Translation

- Lots and lots of data.
- Count co-occurances.
- Normalize.

Constructing Word Embeddings

Distributional Hypothesis (Harris 1954)

Words are determined by the company they keep.

Translation

- Lots and lots of data.
- Count co-occurances.
- Normalize.
- Reduce the dimension.

Constructing Word Embeddings

An Example

He is the king and she is the queen.

Constructing Word Embeddings

An Example

He is the king and she is the queen.

	he	she	is	the	and	queen	king
he	1	0	1	1	0	0	1
she	0	1	1	2	1	1	1
is	1	1	2	2	2	1	1
the	1	2	2	2	2	1	1
and	0	1	2	2	1	0	1
queen	0	1	1	1	0	1	0
king	1	1	2	1	1	0	1

Constructing Word Embeddings

An Example

He is the king *and she is the queen.*

	he	she	is	the	and	queen	king
he	1	0	1	1	0	0	1
she	0	1	1	2	1	1	1
is	1	1	2	2	2	1	1
the	1	2	2	2	2	1	1
and	0	1	2	2	1	0	1
queen	0	1	1	1	0	1	0
king	1	1	2	1	1	0	1

Constructing Word Embeddings

An Example

*He is **the king and she is the queen**.*

	he	she	is	the	and	queen	king
he	1	0	1	1	0	0	1
she	0	1	1	2	1	1	1
is	1	1	2	2	2	1	1
the	1	2	2	2	2	1	1
and	0	1	2	2	1	0	1
queen	0	1	1	1	0	1	0
king	1	1	2	1	1	0	1

Constructing Word Embeddings

An Example, Normalized

He is the king and she is the queen.

	he	she	is	the	and	queen	king
he	0.25	0.00	0.09	0.09	0.00	0.00	0.17
she	0.00	0.14	0.09	0.18	0.14	0.25	0.17
is	0.25	0.14	0.18	0.18	0.29	0.25	0.17
the	0.25	0.29	0.18	0.18	0.29	0.25	0.17
and	0.00	0.14	0.18	0.18	0.14	0.00	0.17
queen	0.00	0.14	0.09	0.09	0.00	0.25	0.00
king	0.25	0.14	0.18	0.09	0.14	0.00	0.17

Constructing Word Embeddings

An Example, with Lower Dimensions!

He is the king and she is the queen.

	?	?	?
`he`	-0.22	0.21	-0.04
`she`	-0.37	-0.16	0.05
`is`	-0.56	0.00	-0.07
`the`	-0.61	-0.03	-0.05
`and`	-0.30	0.02	0.20
`queen`	-0.21	-0.20	-0.05
`king`	-0.37	0.18	0.02

Constucting Word Embeddings

Methods and Tools

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)
  - Simpler mathematically.

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)
  - Simpler mathematically.
  - Based on the ideas and insights of Mikolov et al.

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)
  - Simpler mathematically.
  - Based on the ideas and insights of Mikolov et al.
- FASTTEXT (Bojanowski et al. 2017)

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)
  - Simpler mathematically.
  - Based on the ideas and insights of Mikolov et al.
- FASTTEXT (Bojanowski et al. 2017)
  - WORD2VEC with subword information.

Constucting Word Embeddings

Methods and Tools

- Latent Semantic Analysis (LSA)
  - SVD
  - Computationally expensive!
- WORD2VEC (Mikolov et al. 2013)
  - Fast and cheap!
  - Popularized word embeddings and analogies.
- GLOVE (Pennington et al. 2014)
  - Simpler mathematically.
  - Based on the ideas and insights of Mikolov et al.
- FASTTEXT (Bojanowski et al. 2017)
  - WORD2VEC with subword information.
  - Works well for languages with more inflection than English (e.g German)

Outline

Computing Similarity

Back to the Distributional Hypothesis

- Similar contexts $\Rightarrow$ similar meaning

Computing Similarity

Back to the Distributional Hypothesis

- Similar contexts $\Rightarrow$ similar meaning
- In word embeddings?

Computing Similarity

Back to the Distributional Hypothesis

- Similar contexts $\Rightarrow$ similar meaning
- In word embeddings?
- Similar meaning $\Rightarrow$ close vectors

Computing Similarity

Cosine Similarity

$$\cos(x, y) = \cos\theta = \frac{\langle x, y \rangle}{\|x\|\|y\|}$$

Computing Similarity

Cosine Similarity

$$\cos(x, y) = \cos\theta = \frac{\langle x, y \rangle}{\|x\| \|y\|}$$

For normalized vectors:

$$\cos(x, y) = \langle x, y \rangle$$

Computing Similarity

In our example

He is the king and she is the queen.

Computing Similarity

In our example

He is the king and she is the queen.

- $\cos(\texttt{he}, \texttt{king}) = 0.95$
- $\cos(\texttt{he}, \texttt{queen}) = 0.10$
- $\cos(\texttt{she}, \texttt{king}) = 0.90$
- $\cos(\texttt{she}, \texttt{queen}) = 0.63$

Computing Similarity

In our example

He *is the king and she is the* ***queen***.

- $\cos(\texttt{he}, \texttt{king}) = 0.95$
- $\cos(\texttt{he}, \texttt{queen}) = \mathbf{0.10}$
- $\cos(\texttt{she}, \texttt{king}) = 0.90$
- $\cos(\texttt{she}, \texttt{queen}) = 0.63$

Outline

Analogies

"a is to a^* as b is to b^*", common notation:

$$a : a^* :: b : b^*$$

Analogies

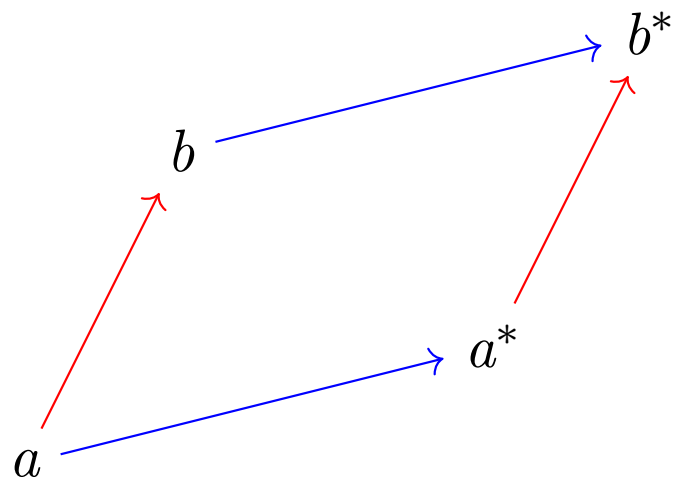
"a is to a^* as b is to b^*", common notation:

$$a : a^* :: b : b^*$$

Mikolov et al. realized that

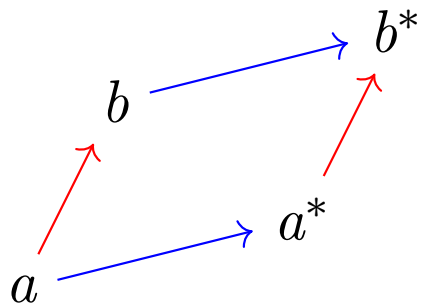
$$a^* - a = b^* - b$$

holds. This is a parallelogram!



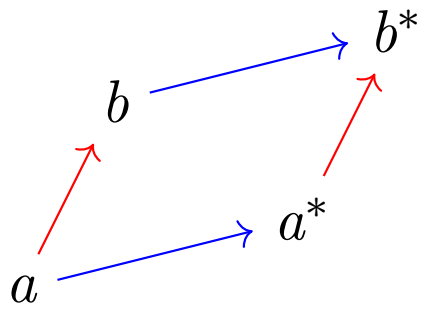
Symmetries

$$a^* - a = b^* - b$$

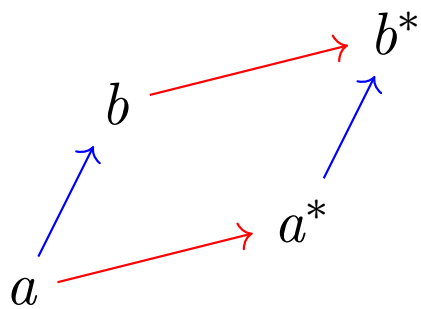


Symmetries

$$a^* - a = b^* - b$$

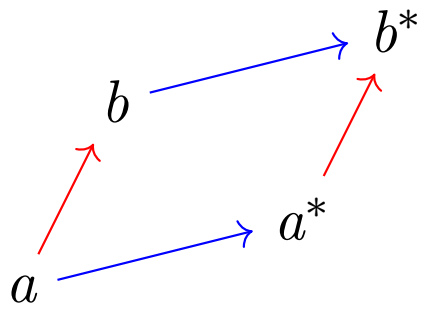


$$b - a = b^* - a^*$$

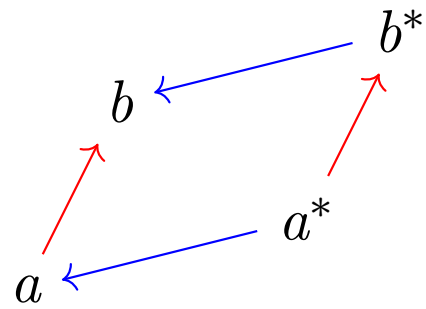


Symmetries

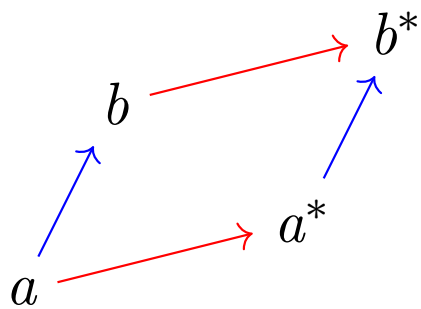
$a^* - a = b^* - b$



$a - a^* = b - b^*$

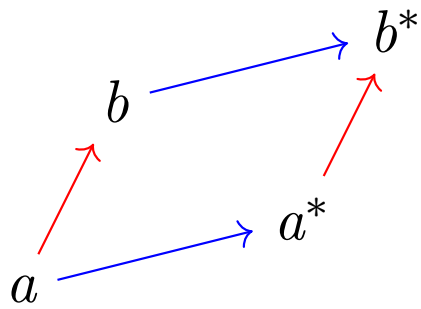


$b - a = b^* - a^*$

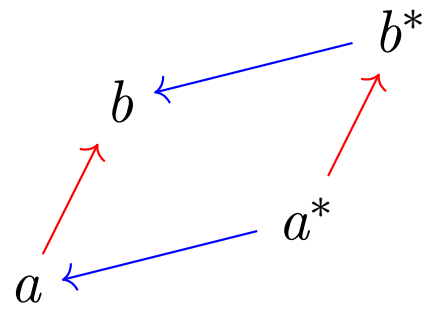


Symmetries

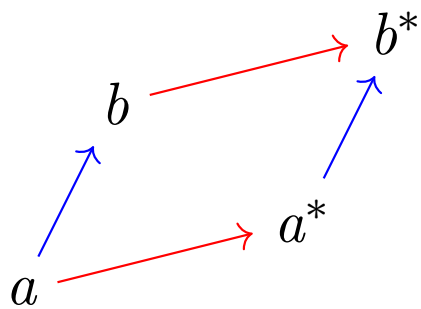
$a^* - a = b^* - b$



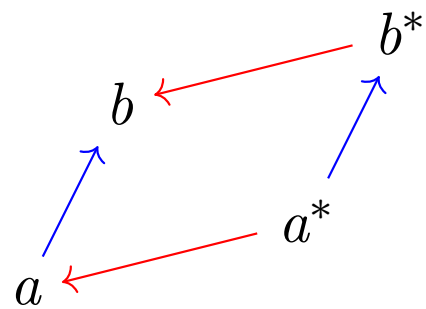
$a - a^* = b - b^*$



$b - a = b^* - a^*$



$a - b = a^* - b^*$



Symmetries

$$
\begin{array}{ll}
a : a^* :: b : b^* & a : b :: a^* : b^* \\
a^* : a :: b^* : b & a^* : b^* :: a : b \\
b : b^* :: a : a^* & b : a :: b^* : a^* \\
b^* : b :: a^* : a & b^* : a^* :: b : a
\end{array}
$$

Symmetries

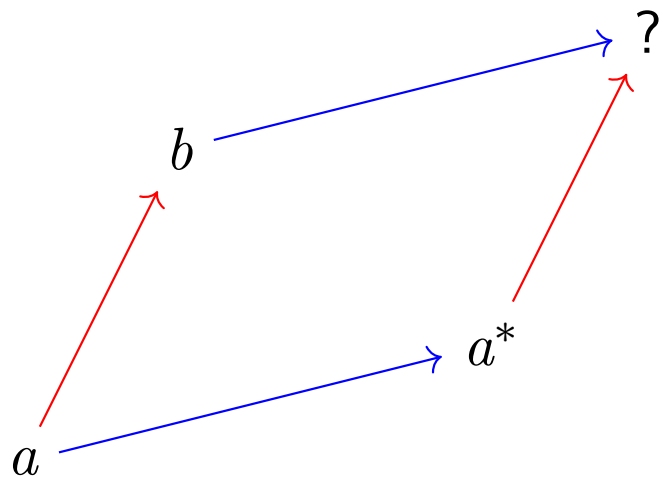
$$
\begin{array}{ll}
a : a^* :: b : b^* & a : b :: a^* : b^* \\
a^* : a :: b^* : b & a^* : b^* :: a : b \\
b : b^* :: a : a^* & b : a :: b^* : a^* \\
b^* : b :: a^* : a & b^* : a^* :: b : a
\end{array}
$$

$$
\begin{array}{ll}
\texttt{he} : \texttt{king} :: \texttt{she} : \texttt{queen} & \texttt{he} : \texttt{she} :: \texttt{king} : \texttt{queen} \\
\texttt{king} : \texttt{he} :: \texttt{queen} : \texttt{she} & \texttt{king} : \texttt{queen} :: \texttt{he} : \texttt{she} \\
\texttt{she} : \texttt{queen} :: \texttt{he} : \texttt{king} & \texttt{she} : \texttt{he} :: \texttt{queen} : \texttt{king} \\
\texttt{queen} : \texttt{she} :: \texttt{king} : \texttt{he} & \texttt{queen} : \texttt{king} :: \texttt{she} : \texttt{he}
\end{array}
$$

Outline

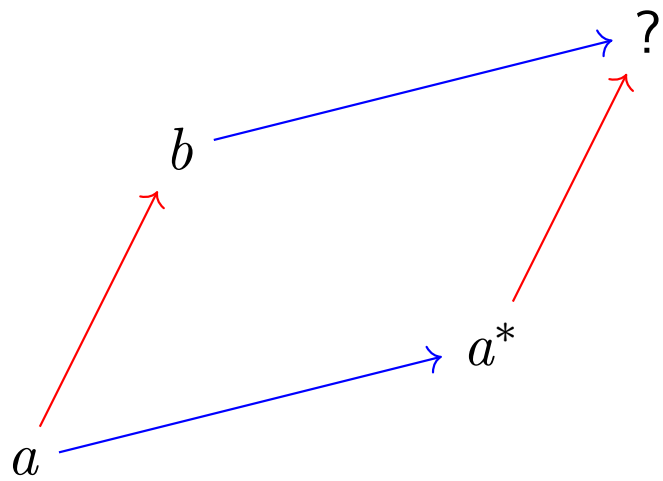
Solving Analogies

Given a, a^* and b, how do we find b^*?



Solving Analogies

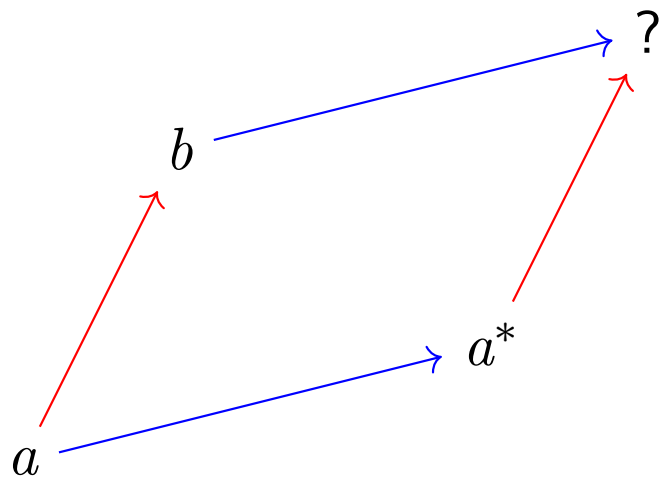
Given a, a^* and b, how do we find b^*?



Using cosine similarity!

Solving Analogies

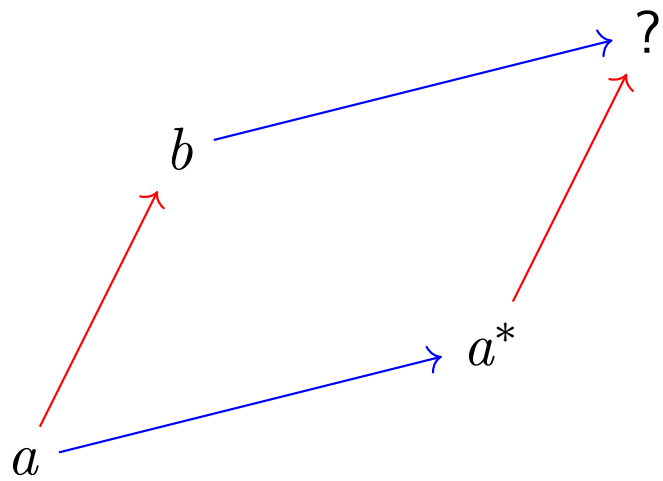
Given a, a^* and b, how do we find b^*?



Using cosine similarity! $a^* - a = b^* - b$, so $b^* = b + a^* - a$.

Solving Analogies

Given a, a^* and b, how do we find b^*?



Using cosine similarity! $a^* - a = b^* - b$, so $b^* = b + a^* - a$.

$$b^* = \text{argmax}_v \cos(v, b + a^* - a)$$

An Equivalence

Expanding the RHS of the previous equation yields

An Equivalence

Expanding the RHS of the previous equation yields

$$b^* = \text{argmax}_v \cos(v, b) + \cos(v, a^*) - \cos(v, a)$$

(called 3CosAdd).

Outline

Understanding 3CosAdd

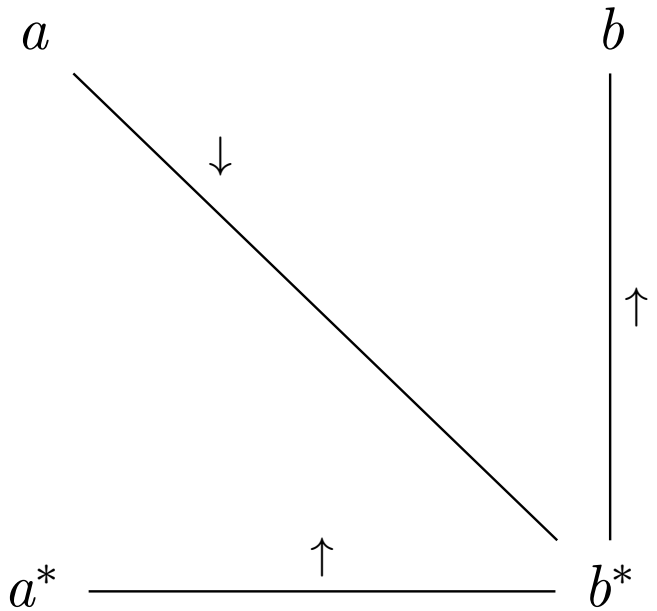
$$b^* = \text{argmax}_v \cos(v, b) + \cos(v, a^*) - \cos(v, a)$$

Find the best b^* that is similar to b and a^* and dissimilar to a.

Understanding 3CosAdd

$$b^* = \text{argmax}_v \cos(v, b) + \cos(v, a^*) - \cos(v, a)$$

Find the best b^* that is similar to b and a^* and dissimilar to a.



Introducing 3CosMul

- Addition is tricky since one term can dominate the other.

Introducing 3CosMul

- Addition is tricky since one term can dominate the other.
- Use multiplication instead!

Introducing 3CosMul

- Addition is tricky since one term can dominate the other.
- Use multiplication instead!

$$b^* = \text{argmax}_v \frac{\cos(v, b)\cos(v, a^*)}{\cos(v, a)}$$

Problem?

- What would be your answer to `he` : `food` :: `she` : x ?

Problem?

- What would be your answer to `he` : `food` :: `she` : x ?
- Solving analogies this way assumes that *the first part of the analogy makes sense!*

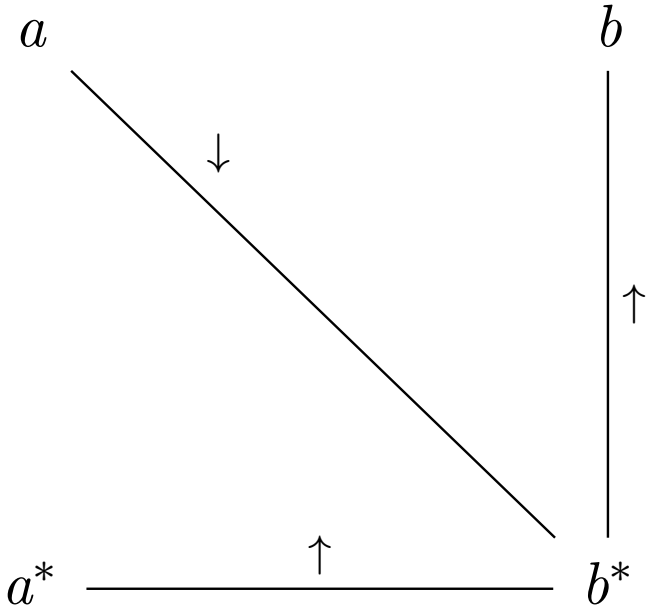
Problem?

- What would be your answer to `he` : `food` :: `she` : x ?
- Solving analogies this way assumes that *the first part of the analogy makes sense!*
- An issue often overlooked!

Problem?

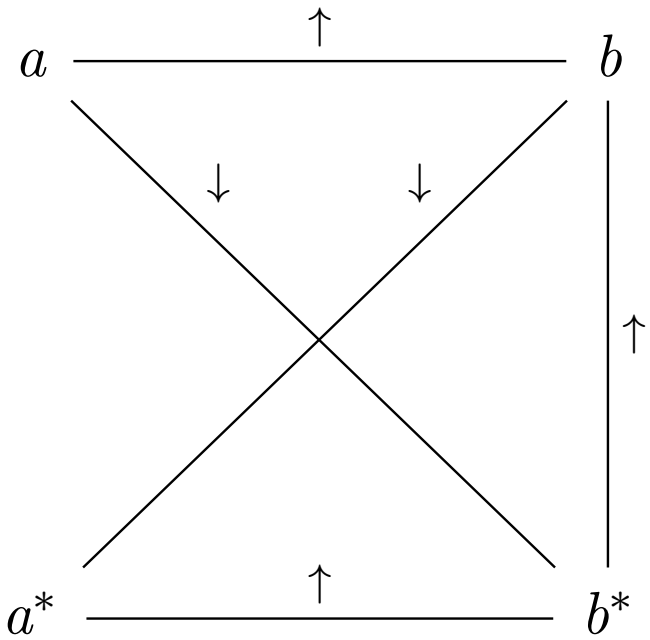
- What would be your answer to `he` : `food` :: `she` : x ?
- Solving analogies this way assumes that *the first part of the analogy makes sense!*
- An issue often overlooked!
- How can we validate the first part?

Symmetries to the Rescue



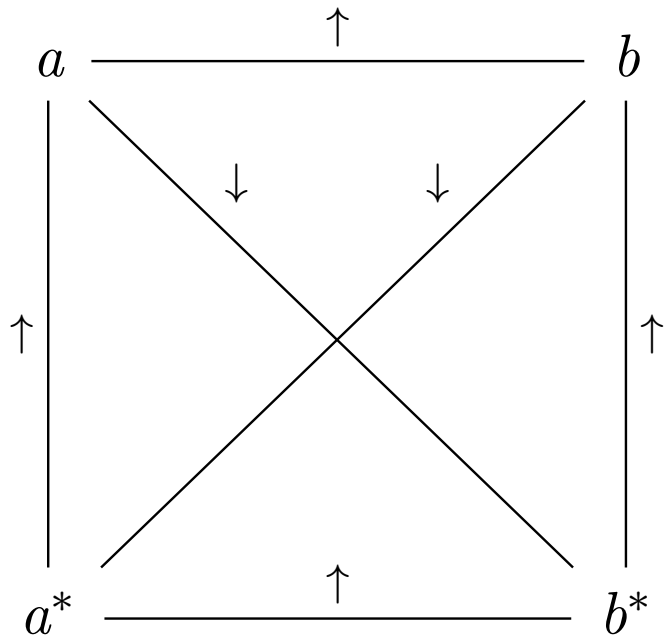
- $a : a^* :: b : b^*$
- $a : b :: a^* : b^*$

Symmetries to the Rescue



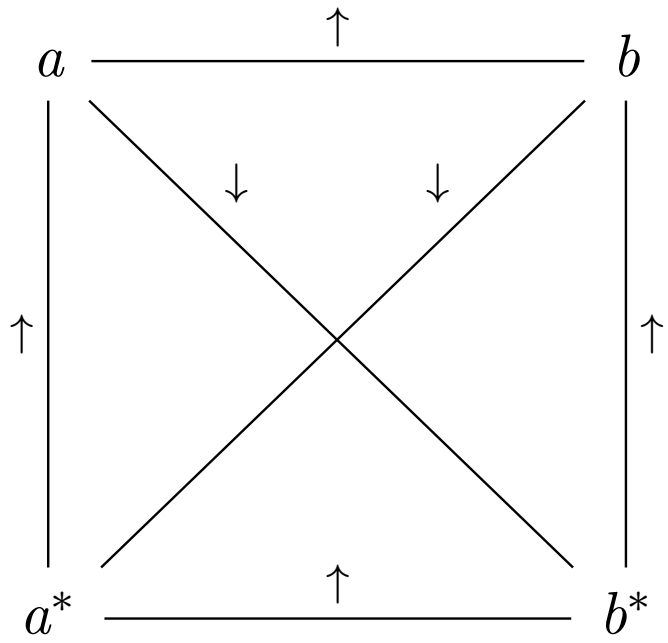
- $a : a^* :: b : b^*$
- $a : b :: a^* : b^*$
- $a^* : a :: b^* : b$
- $a^* : b^* :: a : b$

Symmetries to the Rescue



- $a : a^* :: b : b^*$
- $a : b :: a^* : b^*$
- $a^* : a :: b^* : b$
- $a^* : b^* :: a : b$
- $b : b^* :: a : a^*$
- $b : a :: b^* : a^*$

Symmetries to the Rescue

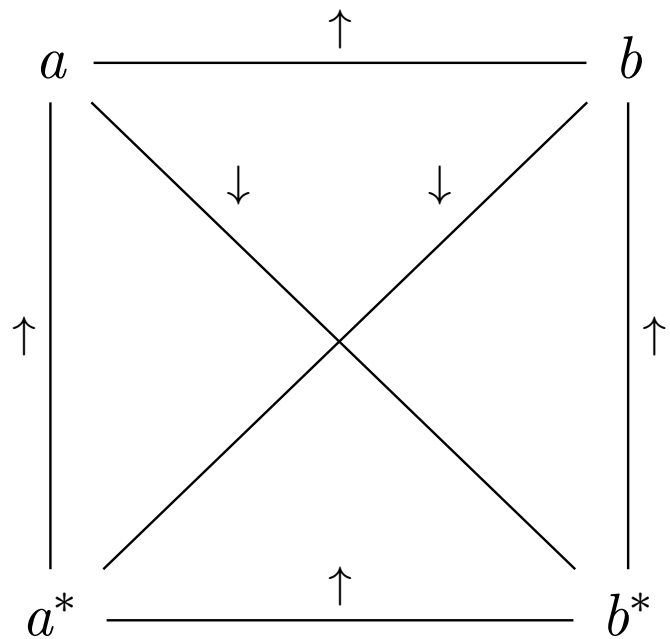


- $a : a^* :: b : b^*$
- $a : b :: a^* : b^*$
- $a^* : a :: b^* : b$
- $a^* : b^* :: a : b$
- $b : b^* :: a : a^*$
- $b : a :: b^* : a^*$
- $b^* : b :: a^* : a$
- $b^* : a^* :: b : a$

Symmetries to the Rescue

Translating this into an expression:

$$S(a, a^*, b, b^*) = \frac{\cos(a, b)\cos(a, a^*)\cos(b, b^*)\cos(a^*, b^*)}{\cos(a, b^*)\cos(b, a^*)}$$



Some Properties

- S generalizes 3CosMul!

$$\operatorname{argmax}_v S(a, a^*, b, v) = \operatorname{argmax}_v \frac{\cos(b, v)\cos(a^*, v)}{\cos(a, v)}$$

Some Properties

- S generalizes $3\textsc{CosMul}$!

$$\text{argmax}_v\, S(a, a^*, b, v) = \text{argmax}_v \frac{\cos(b, v)\cos(a^*, v)}{\cos(a, v)}$$

- $S(a, a^*, a, a^*) = 1$ (good analogies should score close to this)

Some Properties

- S generalizes $3\textsc{CosMul}$!

$$\text{argmax}_v\, S(a, a^*, b, v) = \text{argmax}_v \frac{\cos(b, v)\cos(a^*, v)}{\cos(a, v)}$$

- $S(a, a^*, a, a^*) = 1$ (good analogies should score close to this)
- $S(a, x, b, x) = \cos(a, b)$ (valid analogies should score above this)

Outline

Overview

- Analogies can contain bias.

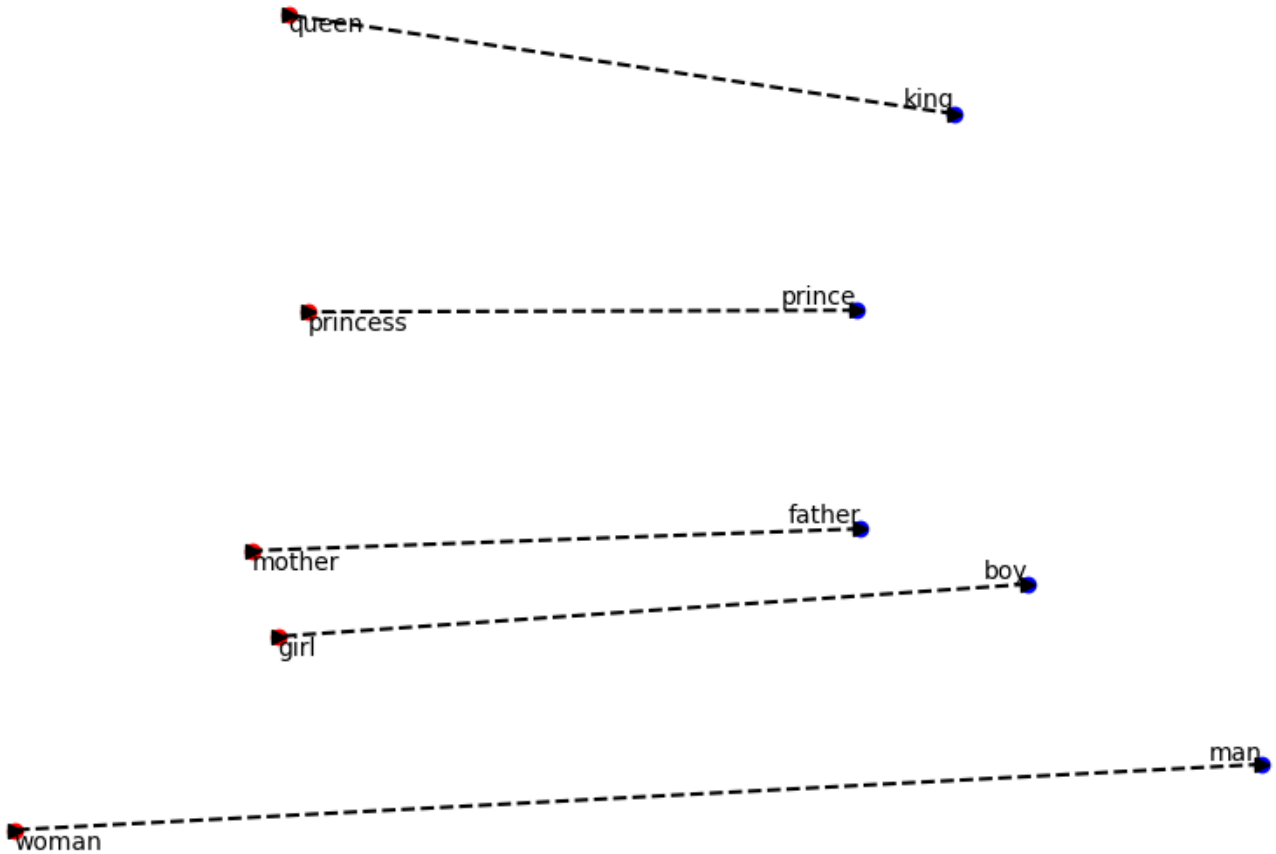
Overview

- Analogies can contain bias.
- Analogies are linear relationships.

Overview

- Analogies can contain bias.
- Analogies are linear relationships.
- $\Rightarrow$ Is bias represented by a bias dimension?

Gender/Royalty Dimensions



Outline

Word Projections

- Pick a type of bias we want to capture (e.g gender)

Word Projections

- Pick a type of bias we want to capture (e.g gender)
- Pick a seed pair that *only differs by this bias* (e.g `he` and `she`)

Word Projections

- Pick a type of bias we want to capture (e.g gender)
- Pick a seed pair that *only differs by this bias* (e.g `he` and `she`)
- Pick some words that we want to evaluate (e.g occupations)

Word Projections

- Pick a type of bias we want to capture (e.g gender)
- Pick a seed pair that *only differs by this bias* (e.g `he` and `she`)
- Pick some words that we want to evaluate (e.g occupations)
- Project the words onto the `he` − `she` line and select the extreme ends.

Word Projections

English

Top Male	Top Female
carpenter	midwife
engineer	nurse
soldier	librarian
surveyor	dancer
blacksmith	housekeeper
mason	teacher
janitor	cook
shoemaker	student
smith	designer
architect	cashier

Word Projections

Swedish

Top Male	Top Female
mekaniker	barnmorska
trumslagare	mannekäng
rörmokare	sjuksköterska
byggare	uska
grävmaskinist	undersköterska
ingenjör	prostituerad
kolare	dietist
biskop	hembiträde
blåsare	kontorist
basist	flygvärdinna

Word Projections

Important Note

- Words that appear together are pulled together in the embedding.

Word Projections

Important Note

- Words that appear together are pulled together in the embedding.
- More women than men are midwives.

Word Projections

Important Note

- Words that appear together are pulled together in the embedding.
- More women than men are midwives.
- $\Rightarrow$ `she` will appear near `midwife` more often than `he`.

Word Projections

Important Note

- Words that appear together are pulled together in the embedding.
- More women than men are midwives.
- $\Rightarrow$ `she` will appear near `midwife` more often than `he`.
- $\Rightarrow$ `she` will be closer to `midwife` in the embedding.

Word Projections

Important Note

- Words that appear together are pulled together in the embedding.
- More women than men are midwives.
- $\Rightarrow$ `she` will appear near `midwife` more often than `he`.
- $\Rightarrow$ `she` will be closer to `midwife` in the embedding.
- Discretization makes it look biased! (Perhaps more than it is!)

Outline

Generating Analogies

- What if we let the embeddings create analogies?

Generating Analogies

- What if we let the embeddings create analogies?
- $S(a, a^*, b, b^*)$ can be used to rank analogies!

Generating Analogies

- What if we let the embeddings create analogies?
- $S(a, a^*, b, b^*)$ can be used to rank analogies!
- An algorithm:

Generating Analogies

- What if we let the embeddings create analogies?
- $S(a, a^*, b, b^*)$ can be used to rank analogies!
- An algorithm:
  1. Start with a seed pair, e.g `he` and `she`.

Generating Analogies

- What if we let the embeddings create analogies?
- $S(a, a^*, b, b^*)$ can be used to rank analogies!
- An algorithm:
  1. Start with a seed pair, e.g `he` and `she`.
  2. Find the best x and y measured by $S(\texttt{he}, x, \texttt{she}, y)$.

Generating Analogies

English

she	he
herself	himself
Her	His
she	he
She	He
her	his
spokeswoman	spokesman
sisters	brothers
woman	man
actress	actor
Ms.	Mr.
niece	nephew

Generating Analogies

Swedish

hon	han
hennes	hans
henne	honom
Hon	Han
hon	han
tjejen	killen
tjej	kille
systrar	bröder
syster	bror
systrarna	bröderna
dotter	son
kvinnan	mannen

Generating Analogies

Some Biased Ones

- `she` : `fabulous` :: `he` : `fantastic`
- `she` : `hair` :: `he` : `beard`
- `she` : `vocalist` :: `he` : `guitarist`

Generating Analogies

Some Biased Ones

- she : fabulous :: he : fantastic
- she : hair :: he : beard
- she : vocalist :: he : guitarist
- hon : klänning :: han : skjorta
- hon : kjol :: han : skjorta

Outline

English vs. Swedish

- Swedes try to make the language gender neutral, e.g *hen* (he/she).

English vs. Swedish

- Swedes try to make the language gender neutral, e.g *hen* (he/she).
- Is English thus more biased than Swedish?

English vs. Swedish

- Swedes try to make the language gender neutral, e.g *hen* (he/she).
- Is English thus more biased than Swedish?
- How can we compare them?

English vs. Swedish

- For all methods, the similarity between the seed words is important.

English vs. Swedish

- For all methods, the similarity between the seed words is important.
- $\cos(x, y) = 1 \;\Rightarrow\; x = y$

English vs. Swedish

- For all methods, the similarity between the seed words is important.
- $\cos(x, y) = 1 \;\Rightarrow\; x = y$
- Similar words $\Rightarrow$ low bias

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - $(\texttt{hon}, \texttt{han})$ and $(\texttt{she}, \texttt{he})$

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - (`hon`, `han`) and (`she`, `he`)
  - (`kvinnor`, `män`) and (`women`, `men`)

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - (`hon`, `han`) and (`she`, `he`)
  - (`kvinnor`, `män`) and (`women`, `men`)
  - etc.

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - (`hon`, `han`) and (`she`, `he`)
  - (`kvinnor`, `män`) and (`women`, `men`)
  - etc.
- Compare the cosine similarities using a sign test.

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - $(\texttt{hon}, \texttt{han})$ and $(\texttt{she}, \texttt{he})$
  - $(\texttt{kvinnor}, \texttt{män})$ and $(\texttt{women}, \texttt{men})$
  - etc.
- Compare the cosine similarities using a sign test.
  - $\cos(\texttt{hon}, \texttt{han}) > \cos(\texttt{she}, \texttt{he})$?

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - $(\texttt{hon}, \texttt{han})$ and $(\texttt{she}, \texttt{he})$
  - $(\texttt{kvinnor}, \texttt{män})$ and $(\texttt{women}, \texttt{men})$
  - etc.
- Compare the cosine similarities using a sign test.
  - $\cos(\texttt{hon}, \texttt{han}) > \cos(\texttt{she}, \texttt{he})$?
  - $\cos(\texttt{kvinnor}, \texttt{män}) > \cos(\texttt{women}, \texttt{men})$?

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - $(\texttt{hon}, \texttt{han})$ and $(\texttt{she}, \texttt{he})$
  - $(\texttt{kvinnor}, \texttt{män})$ and $(\texttt{women}, \texttt{men})$
  - etc.
- Compare the cosine similarities using a sign test.
  - $\cos(\texttt{hon}, \texttt{han}) > \cos(\texttt{she}, \texttt{he})$?
  - $\cos(\texttt{kvinnor}, \texttt{män}) > \cos(\texttt{women}, \texttt{men})$?
  - etc.

Sign Test

Translated into a concrete test:

- Find matching gendered pairs in English and Swedish.
  - $(\texttt{hon}, \texttt{han})$ and $(\texttt{she}, \texttt{he})$
  - $(\texttt{kvinnor}, \texttt{män})$ and $(\texttt{women}, \texttt{men})$
  - etc.
- Compare the cosine similarities using a sign test.
  - $\cos(\texttt{hon}, \texttt{han}) > \cos(\texttt{she}, \texttt{he})$?
  - $\cos(\texttt{kvinnor}, \texttt{män}) > \cos(\texttt{women}, \texttt{men})$?
  - etc.
- Significant result $\Rightarrow$ Swedish less biased than English

Sign Test

Results

- Two pairs of embeddings tested.

Sign Test

Results

- Two pairs of embeddings tested.
- $p = 0.21$ for both pairs.

Sign Test

Results

- Two pairs of embeddings tested.
- $p = 0.21$ for both pairs.
- $p > 0.05$

Sign Test

Results

- Two pairs of embeddings tested.
- $p = 0.21$ for both pairs.
- $p > 0.05$
- Well...

Summary

Conclusions

- Validate analogies with the score S.
- Discretization exaggerates stereotypes, use continuous measures!
- Normal language use makes word embeddings sexist.

Summary

Conclusions

- Validate analogies with the score S.
- Discretization exaggerates stereotypes, use continuous measures!
- Normal language use makes word embeddings sexist.

Closing Thoughts

- Debias embeddings before using them.
- Swedish might still be less biased than English.

Further Reading

The thesis, code and data will be published online when finished.

`https://precenth.eu/word-embeddings.html`