Generating Bags of Words from the Sums of their Word Embeddings

A greedy algorithms for (re-)creating the unordered collection of words from a sum of word embeddings representation

Lyndon White,
Roberto Togneri, Wei Liu, Mohammed Bennamoun

School of Electical, Electronic and Computer Engineering
The University of Western Australia

What are sentence vector representations?

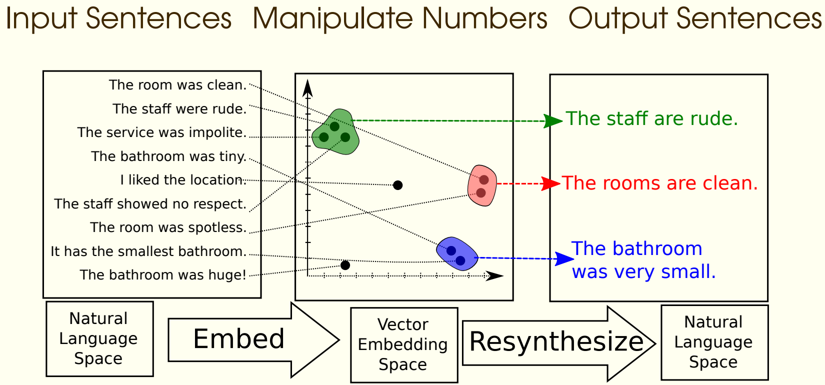
Methods for representing key information about a sentence, as a vector

Classical (Non-compositional): LSA, LDA, BOW ...

Compositional: RAE, RvNN, ...

Noncompositional: PV-DM, PV-DBOW, **SOWE**

We have turned sentences into numeric vectors, now we want to turn them back.



Why are we converting SOWE to BOW?

- Part-way step towards sentence generation.

Why are we converting SOWE to BOW?

- Part-way step towards sentence generation.
- Translating various media to keywords via common vector space.

Why are we converting SOWE to BOW?

- Part-way step towards sentence generation.
- Translating various media to keywords via common vector space.
- Theoretical implications on what information is maintained by the SOWE.

We have turned sentences into numeric vectors, now we want to turn them back.

Sentence: It was the best of times, it was the worst of times

We have turned sentences into numeric vectors, now we want to turn them back.

Sentence: It was the best of times, it was the worst of times

Vector representation: $[0.79, 1.27, 0.28, ..., 1.29]$

We have turned sentences into numeric vectors, now we want to turn them back.

Sentence: It was the best of times, it was the worst of times

Vector representation: $[0.79, 1.27, 0.28, ..., 1.29]$

BOW output: `{best: 1,times: 2, worst: 1,`
`it: 2, of: 2, the: 2, was: 2,, : 1}`

We have turned sentences into numeric vectors, now we want to turn them back.

Sentence: It was the best of times, it was the worst of times

Vector representation: $[0.79, 1.27, 0.28, ..., 1.29]$

BOW output: `{best: 1,times: 2, worst: 1, it: 2, of: 2, the: 2, was: 2,, : 1}`

Sentence: It was the worse of times, it was the best of times

Existing methods do not produce closely matching sentences.

- Iyyer et al's compositional method
- Bowman et al's RNN based method

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

Existing methods do not produce closely matching sentences.

- Iyyer et al's compositional method
- Bowman et al's RNN based method

- Both are demonstrated to produce loosely similar sentences.

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

Existing methods do not produce closely matching sentences.

- Iyyer et al's compositional method
- Bowman et al's RNN based method

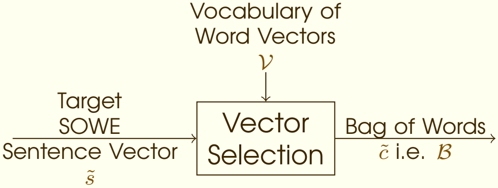
- Both are demonstrated to produce loosely similar sentences.
- Neither has show a demonstration on any large scale corpus.

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

We solve the objective function to get a bag of words.

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$



We solve the objective function to get a bag of words.

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

We solve the objective function to get a bag of words.

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

Input Vector $\tilde{s} = [0.79, 1.27, 0.28, ..., 1.29]$

We solve the objective function to get a bag of words.

Find the inclusion vector $\tilde{c} = [c_1, c_2, ... c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

Input Vector $\tilde{s} = [0.79, 1.27, 0.28, ..., 1.29]$

Vector Selection

$$
\tilde{s} \approx \begin{array}{rc} & 1 \times [0.19, 0.50, 0.14, ..., 0.59] \\ + & 2 \times [-0.15, 0.19, 0.03, ..., -0.17] \\ + & ... \\ + & 0 \times [0.19, 2.10, 1.34, ..., 1.20] \\ + & 1 \times [0.79, 1.27, 0.28, ..., 1.29] \end{array}
$$

We solve the objective function to get a bag of words.

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

Input Vector $\tilde{s} = [0.79, 1.27, 0.28, ..., 1.29]$

Vector Selection

$$\tilde{s} \approx \begin{array}{rl} & 1 \times [0.19, 0.50, 0.14, ..., 0.59] \\ + & 2 \times [-0.15, 0.19, 0.03, ..., -0.17] \\ + & \quad ... \\ + & 0 \times [0.19, 2.10, 1.34, ..., 1.20] \\ + & 1 \times [0.79, 1.27, 0.28, ..., 1.29] \end{array}$$

BOW `{best: 1,times: 2, worst: 1, it: 2, of: 2, the: 2, was: 2,, : 1}`

How to solve the objective function? Greedily

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

- Similarities to Knapsack family of problems.
  - Provably NP-Hard

How to solve the objective function? Greedily

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

- Similarities to Knapsack family of problems.
  - Provably NP-Hard
- Very high dimensionality of inclusion vector
  - $|\mathcal{V}| \approx 40,000$ for Brown Corpus
  - $|\mathcal{V}| \approx 180,000$ for Books Corpus

How to solve the objective function? Greedily

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \ \tilde{x}_j \, c_j)$

- Similarities to Knapsack family of problems.
  - Provably NP-Hard
- Very high dimensionality of inclusion vector
  - $|\mathcal{V}| \approx 40,000$ for Brown Corpus
  - $|\mathcal{V}| \approx 180,000$ for Books Corpus
- A greedy algorithm is linear time in n

A more direct bag notation for vector selection problem.

Rather than writing:

Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

A more direct bag notation for vector selection problem.

Rather than writing:
Find the inclusion vector $\tilde{c} = [c_1, c_2, ...c_{|\mathcal{V}|}] \in \mathbb{N}_0^{|\mathcal{V}|}$ that for we have $\min d(\tilde{s}, \sum_{\tilde{x}_j \in \mathcal{V}} \tilde{x}_j \, c_j)$

We can equivalently say:
Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Greedy Addition: where you add the best vector to your current bag, and repeat.

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

1. For each vector $\tilde{x}_j$ in the vocabulary consider $d(\tilde{s}, \Sigma(\mathcal{B}) + \tilde{x}_j)$
2. Add in the vector that gets the total closest to $\tilde{s}$ $\mathcal{B} \leftarrow \mathcal{B} \cup \{\tilde{x}_\star\}$
   - unless adding nothing would be better – then terminate
3. Repeat

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\quad \tilde{s} = 148 \quad d(x, y) = |x - y|$

1. $\mathcal{B} = [\,]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\tilde{s} = 148$ $d(x, y) = |x - y|$

1. $\mathcal{B} = [\,]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$
2. $\mathcal{B} = [100]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 100| = 48$

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\qquad \tilde{s} = 148 \qquad d(x, y) = |x - y|$

1. $\mathcal{B} = [\,]$ $\qquad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$
2. $\mathcal{B} = [100]$ $\qquad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 100| = 48$
3. $\mathcal{B} = [100, 25]$ $\qquad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25)| = 23$

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\tilde{s} = 148$ $d(x, y) = |x - y|$

1. $\mathcal{B} = [\,]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$
2. $\mathcal{B} = [100]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 100| = 48$
3. $\mathcal{B} = [100, 25]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25)| = 23$
4. $\mathcal{B} = [100, 25, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\tilde{s} = 148$ $d(x,y) = |x - y|$

1. $\mathcal{B} = [\,]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$
2. $\mathcal{B} = [100]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 100| = 48$
3. $\mathcal{B} = [100, 25]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25)| = 23$
4. $\mathcal{B} = [100, 25, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
5. $\mathcal{B} = [100, 25, 24]$ No improvement possible $d(\tilde{s}, \Sigma(\mathcal{B})) = 1$

A 1 dimentional example of greedy additon

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\quad \tilde{s} = 148 \quad d(x, y) = |x - y|$

1. $\mathcal{B} = [\,]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 0| = 148$
2. $\mathcal{B} = [100]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - 100| = 48$
3. $\mathcal{B} = [100, 25]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25)| = 23$
4. $\mathcal{B} = [100, 25, 24]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
5. $\mathcal{B} = [100, 25, 24]$ No improvement possible $d(\tilde{s}, \Sigma(\mathcal{B})) = 1$

Fell for greedy trap

1-Subsitution: Lessen the greed by reconsidering past choices

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

1. Consider each word vector in the current bag $\tilde{x}_a \in \mathcal{B}$
2. Would deleting it improve the score?
   $d(\tilde{s}, \Sigma(\mathcal{B}) - \tilde{x}_a) < d(\tilde{s}, \Sigma(\mathcal{B}))$?
3. Can it be swapped for another word to improve the score? $\exists \tilde{x}_b \in \mathcal{V}$ such that
   $d(\tilde{s}, \Sigma(\mathcal{B}) - \tilde{x}_a + \tilde{x}_b)) < d(\tilde{s}, \Sigma(\mathcal{B}))$?

A 1 dimentional example of 1-substitution

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\quad \tilde{s} = 148 \quad d(x, y) = |x - y|$

1. $\mathcal{B} = [100, 25, 24] \quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$

A 1 dimentional example of 1-substitution

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\quad \tilde{s} = 148$ $\quad d(x, y) = |x - y|$

1. $\mathcal{B} = [100, 25, 24]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
2. Consider swapping 100

A 1 dimentional example of 1-substitution

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\quad \tilde{s} = 148$ $\quad d(x, y) = |x - y|$

1. $\mathcal{B} = [100, 25, 24]$ $\quad d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
2. Consider swapping 100
3. Consider swapping 25

A 1 dimentional example of 1-substitution

Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\tilde{s} = 148$ $d(x, y) = |x - y|$

1. $\mathcal{B} = [100, 25, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
2. Consider swapping 100
3. Consider swapping 25
4. $\mathcal{B} = [100, 24, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 24 + 24)| = 0$

A 1 dimentional example of 1-substitution

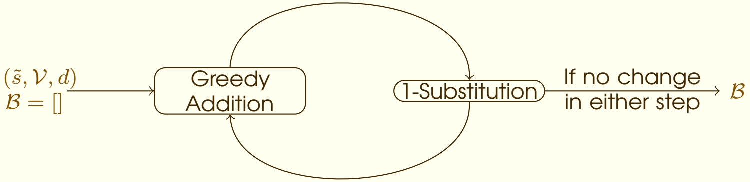
Find the bag of vectors $\mathcal{B}$ (a multi-subset of $\mathcal{V}$), such that we have $\min d(\tilde{s}, \sum_{\tilde{x}_a \in \mathcal{B}} \tilde{x}_a)$

Consider $\mathcal{V} = \{24, 25, 100\}$ $\tilde{s} = 148$ $d(x, y) = |x - y|$

1. $\mathcal{B} = [100, 25, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 25 + 24)| = 1$
2. Consider swapping 100
3. Consider swapping 25
4. $\mathcal{B} = [100, 24, 24]$ $d(\tilde{s}, \Sigma(\mathcal{B})) = |148 - (100 + 24 + 24)| = 0$

Fixed, but there are deeper greedy traps, that can be constructed.

Run until converance



Experimental Setup

Preprocess corpora to only use known words.

- For word embeddings, we use pretrained GloVe

J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014, pp. 1532–1543.

Preprocess corpora to only use known words.

- For word embeddings, we use pretrained GloVe
- Restrict Vector vocab to only words used in corpora
- Preprocess Corpora to remove sentences with words not found in vocabulary.

J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, 2014, pp. 1532–1543.

We used the Brown, and the Books Corpus as generation targets.

W. N. Francis and H. Kucera, "Brown corpus manual," *Brown University*, 1979.

Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *ArXiv preprint arXiv:1506.06724*, 2015.

We used the Brown, and the Books Corpus as generation targets.

Brown Corpus

- Extracts from 500 varied works from 1961
- 40,485 unique words
- 42,004 sentences
- Sentence Length Q3: 25 words

W. N. Francis and H. Kucera, "Brown corpus manual," *Brown University*, 1979.

Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *ArXiv preprint arXiv:1506.06724*, 2015.

We used the Brown, and the Books Corpus as generation targets.

Brown Corpus

- Extracts from 500 varied works from 1961
- 40,485 unique words
- 42,004 sentences
- Sentence Length Q3: 25 words

Books Corpus

- 11,038 unpublished novels
  we use a random subset
- 178,694 unique words
- 66,464 sentences
- Sentence Length Q3: 17 words

W. N. Francis and H. Kucera, "Brown corpus manual," *Brown University*, 1979.

Y. Zhu, R. Kiros, R. Zemel, R. Salakhutdinov, R. Urtasun, A. Torralba, and S. Fidler, "Aligning books and movies: Towards story-like visual explanations by watching movies and reading books," in *ArXiv preprint arXiv:1506.06724*, 2015.

Results

A pair of short example

Sentence: we looked out at the setting sun .
Target BOW: . at looked out setting sun the we
Output BOW: . at looked out setting sun the we

Sentence: i went to the kitchen .
Target BOW: . i kitchen the to went
Output BOW: . i kitchen the to went

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

A short example where the method fails

Sentence: how are you doing ?

Target BOW: ? are doing how you

Output BOW: ? **'re** ~~are~~ **do** ~~doing~~ how **well** ~~you~~

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

A medium length example

Sentence: this is the basis of a comedy of manners first performed in 1892

Target BOW: 1892 a basis comedy first in is manners of of performed the this

Output BOW: 1892 a basis comedy first in is manners of of performed the this

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

A long example

Sentence: thus she leaves her husband and child for aleksei vronsky but all ends sadly when she leaps in front of a train

Target BOW: a aleksei all and but child ends for front her husband in leaps leaves of sadly she she thus train vronsky when

Output BOW: a aleksei all and but child ends for front her husband in leaps leaves of sadly she she thus train vronsky when

---

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

A long example where the method fails.

Sentence: ralph waldo emerson dismissed this poet as the jingle man and james russell lowell called him three-fifths genius and two-fifths sheer fudge

Target BOW: and and as called dismissed emerson fudge genius him james jingle lowell man poet ralph russell sheer the this three-fifths two-fifths waldo

Output BOW: **2008 _..._(13) _..._(34) _..._(44) " aldrick** and and ~~as~~ **both** called dismissed emerson fudge genius **hapless** him **hirsute** james jingle **known** lowell man poet ralph russell sheer the this three-fifths two-fifths waldo **was**

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

Yet another example

Sentence: in a third novel a sailor abandons the patna and meets marlow who in another novel meets kurtz in the congo

Target BOW: a a abandons and another congo in in in kurtz marlow meets meets novel novel patna sailor the the third who

Output BOW: a a abandons and another congo in in in kurtz marlow meets meets novel novel patna sailor the the third who

---

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

A final example

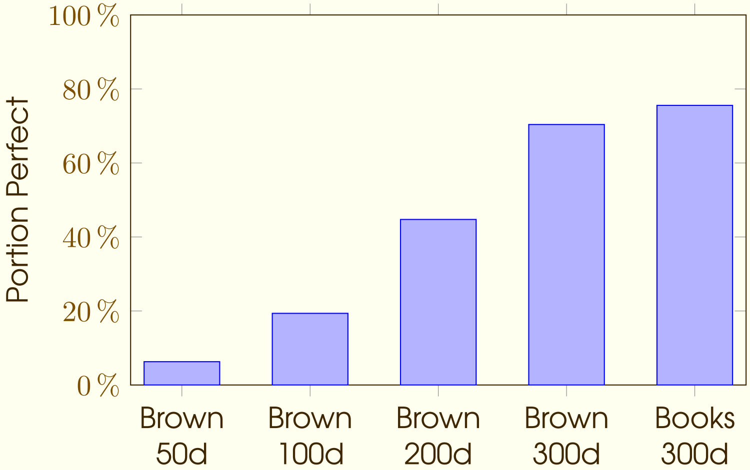
Sentence: name this 1922 novel about leopold bloom written by james joyce

Target BOW: 1922 about bloom by james joyce leopold name novel this written

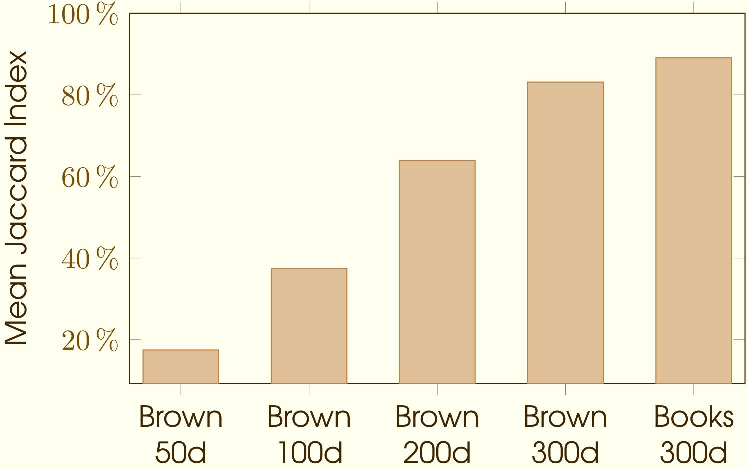
Output BOW: 1922 about bloom by james joyce leopold name novel this written

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

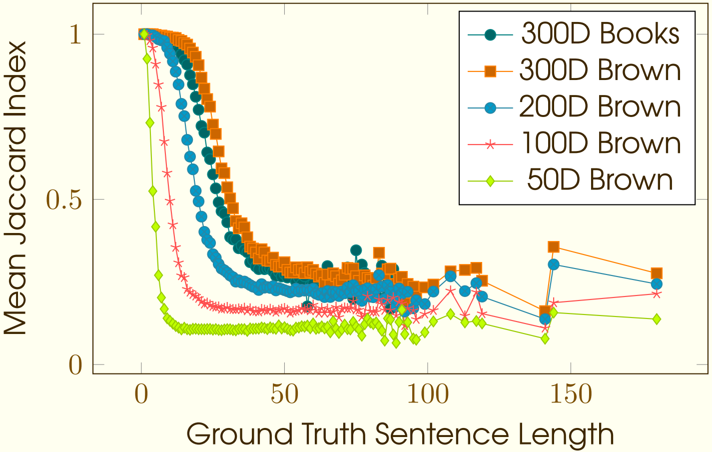
Results: The more dimentions used in the word embeddings, the better the recovery.



Results: The more dimentions used in the word embeddings, the better the recovery.



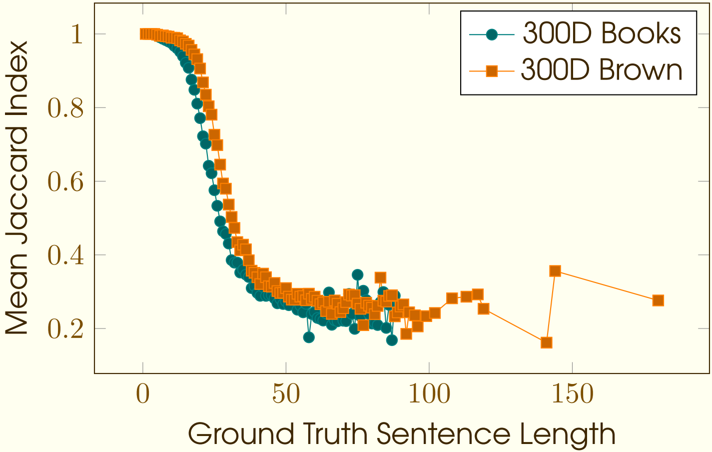
Result: The longer the sentence, the worse recovery



Brown Q3: 25 words

Books Q3: 17 words

Result: The larger the vocabulary, the worse recovery



Brown $|\mathcal{V}| \approx 40,000$

Books $|\mathcal{V}| \approx 180,000$

Conclusion

Future Work: we could to order them to get a sentence.

- Use a language model to find probability of any given sequence.

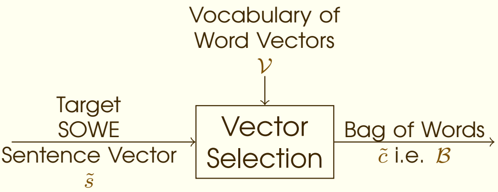
Future Work: we could to order them to get a sentence.

- Use a language model to find probability of any given sequence.
- Not guaranteed to find a single unique order.

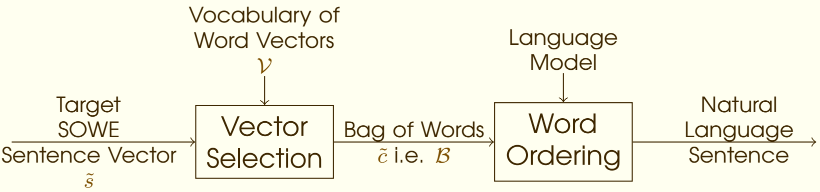
Future Work: we could to order them to get a sentence.

- Use a language model to find probability of any given sequence.
- Not guaranteed to find a single unique order.
- Also NP-hard.

A two step method for generating sentences.



A two step method for generating sentences.



Conclusion: We can often successfully recover the BOW, from the SOWE

- Vector selection with a greedy algorithm
  - This is a broad generalisation of Knapsack Problem
  - Input: SOWE vector
  - Greedy Addition + 1-Substitution til convergence.
  - Output: BOW
- Future work: order the words using a language model.

Appendix

Generating Bags of Words from the Sums of their Word Embeddings

A greedy algorithms for (re-)creating the unordered collection of words from a sum of word embeddings representation

Lyndon White,
Roberto Togneri, Wei Liu, Mohammed Bennamoun

School of Electical, Electronic and Computer Engineering
The University of Western Australia

Recent results suggest sum of word embeddings captures surprising amounts of semantic information

Category	Example
Adhesion to Vertical Surface	There is a magnet on the refrigerator.
Support by Horizontal Surface	There is an apple on the refrigerator.
Support from Above	There is an apple on the branch.
Full Containment	There is an apple in the refrigerator.
Partial Containment	There is an apple in the water.

- Categorise sentences based on the positional component of their meaning.
- Ritter et. al. found sum of word embeddings to outperform all more complex models.

S. Ritter, C. Long, D. Paperno, M. Baroni, M. Botvinick, and A. Goldberg, "Leveraging preposition ambiguity to assess compositional distributional models of semantics," *The Fourth Joint Conference on Lexical and Computational Semantics*, 2015.

Recent results suggest sum of word embeddings captures surprising amounts of semantic information

- We groups MSRP and Opinosis sentences by semantic equivalence forming classes of paraphrases.
- Then used various sentence embeddings as input to a linear SVM to try and classify back into the groups.
- SOWE was amongst top contenders (<0.6% worse than best in both cases)

L. White, R. Togneri, W. Liu, and M. Bennamoun, "How well sentence embeddings capture meaning," in *Proceedings of the 20th Australasian Document Computing Symposium*, ser. ADCS '15, Parramatta, NSW, Australia: ACM, 2015, 9:1–9:8, ISBN: 978-1-4503-4040-3. DOI: 10.1145/2838931.2838932.

Iyyer et al's compositional sentence generation method.

- Variation on the URAE
- Reuses a neural network to (merge up the dependency tree
- Similar to unfold.
- Requires structure of output to be given as a input.

M. Iyyer, J. Boyd-Graber, and H. D. III, "Generating sentences from semantic vector space representations," in *NIPS Workshop on Learning Semantics*, 2014.

Bowman et al's RNN based sentence generation method.

- Use LTSM RNN for decode/encoding step
- Use VAE as representation of posterior probabilities.
- lots of interesting properties and other uses.

S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

Results

Corpus	Word Embedding Dimensions	Portion Perfect	Mean Jaccard Score	Mean Precision	Mean Recall	Mean F1 Score
Brown	50	6.3%	0.175	0.242	0.274	0.265
Brown	100	19.4%	0.374	0.440	0.530	0.477
Brown	200	44.7%	0.639	0.695	0.753	0.720
Brown	300	70.4%	0.831	0.864	0.891	0.876
Books	300	75.6%	0.891	0.912	0.937	0.923