Linguistic Regularities in Continuous Space Word Representations

MIKOLOV, YIH & ZWEIG (2013)

ANDREW WATSON

Outline

- Concepts
  - Continuous Space Word Representations
  - Linguistic Regularities
  - Recurrent Neural Networks
- Experiments
  - Syntactic Test
  - Semantic Test
  - Model - Vector Offset Method
  - Results
- Conclusions

Continuous Space Word Representations

Use of Neural Network Language Models to predict the next word given the previous words

- Recurrent Neural Networks in particular.

Training a neural network language model also provides implicitly learned word representations.

- High dimensional, real valued vectors.

These capture meaningful syntactic and semantic regularities

Capturing Linguistic Regularities

Constant vector offsets between the learned representations of syntactically/semantically words

$$x_{apple} - x_{apples} \approx x_{car} - x_{cars}$$

$$x_{family} - x_{families} \approx x_{car} - x_{cars}$$

$$x_b - x_a \approx x_d - x_c$$

Recurrent Neural Network Language Model

Input: “One-of-N”/”One hot” vector
Dimensionality = Vocabulary Size

$$\mathbf{s}(t) = f\left(\mathbf{U}\mathbf{w}(t) + \mathbf{W}\mathbf{s}(t-1)\right)$$

$$\mathbf{y}(t) = g\left(\mathbf{V}\mathbf{s}(t)\right),$$

$$f(z) = \frac{1}{1+e^{-z}}, \quad g(z_m) = \frac{e^{z_m}}{\sum_k e^{z_k}}.$$

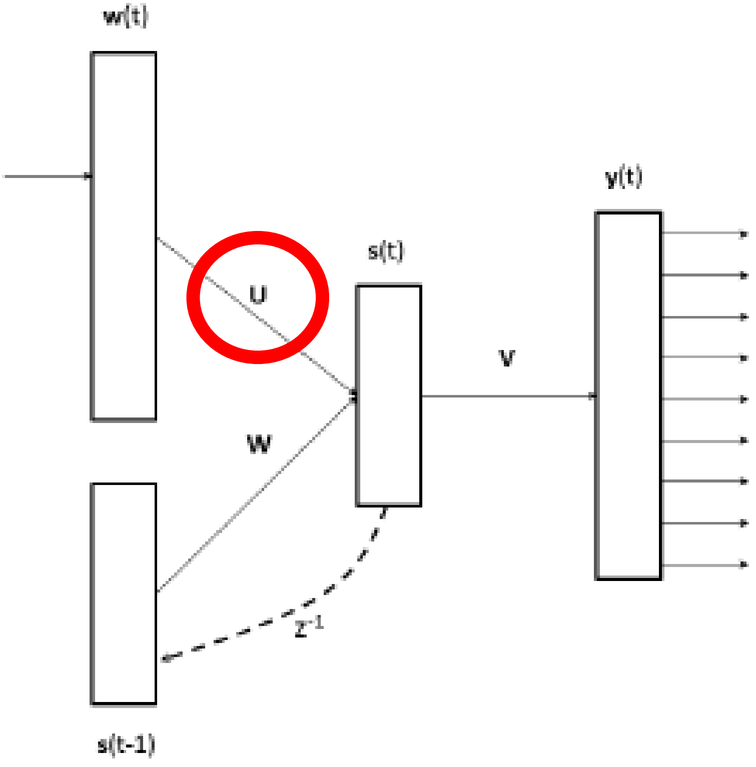


Figure 1: Recurrent Neural Network Language Model.

Testing Regularities

Testing Syntactic Regularity

Test set of analogy questions of the form:

- “A is to B as C is to D”

“car is to cars as bank is to ___”

Tagged 267m words of newspaper text from Penn Treebank POS tags.

Category	Relation	Patterns Tested	# Questions	Example
Adjectives	Base/Comparative	JJ/JJR, JJR/JJ	1000	good:better rough:__
Adjectives	Base/Superlative	JJ/JJS, JJS/JJ	1000	good:best rough:__
Adjectives	Comparative/ Superlative	JJS/JJR, JJR/JJS	1000	better:best rougher:__
Nouns	Singular/Plural	NN/NNS, NNS/NN	1000	year:years law:__
Nouns	Non-possessive/ Possessive	NN/NN_POS, NN_POS/NN	1000	city:city’s bank:__
Verbs	Base/Past	VB/VBD, VBD/VB	1000	see:saw return:__
Verbs	Base/3rd Person Singular Present	VB/VBZ, VBZ/VB	1000	see:sees return:__
Verbs	Past/3rd Person Singular Present	VBD/VBZ, VBZ/VBD	1000	saw:sees returned:__

Table 1: Test set patterns. For a given pattern and word-pair, both orderings occur in the test set. For example, if “see:saw return:__” occurs, so will “saw:see returned:__”.

Testing Semantic Regularity

SemEval-2012 Task 2, *Measuring Relation Similarity*

Given a group of word pairs, order them according to the extent they capture the same relationship as a given reference pair

eg;

“clothing is to shirt as dish is to bowl”

Testing the Model – Vector Offset Method

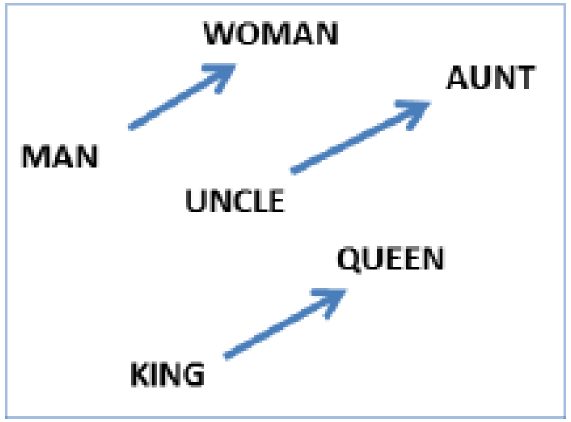
Assume relations are present as vector offsets

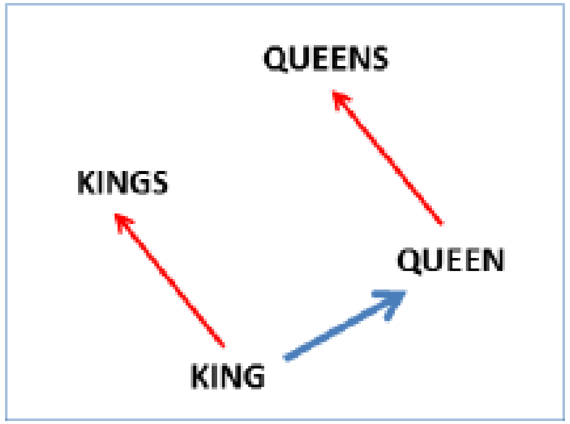
Syntactic Relations

- $y = x_b - x_a + x_c$ Vector Offset
- $w* = argmax_w \frac{x_w y}{||x_w||||y||}$ Cosine Similarity

Semantic Relations

- Word d is given – Difference in cosine similarity between the pairs.





Results

Syntactic Regularities

Method	Adjectives	Nouns	Verbs	All
LSA-80	9.2	11.1	17.4	12.8
LSA-320	11.3	18.1	20.7	16.5
LSA-640	9.6	10.1	13.8	11.3
RNN-80	9.3	5.2	30.4	16.2
RNN-320	18.2	19.0	45.0	28.5
RNN-640	21.0	25.2	54.8	34.7
RNN-1600	**23.9**	**29.2**	**62.2**	**39.6**

Table 2: Results for identifying syntactic regularities for different word representations. Percent correct.

Results

Syntactic Regularities

Method	Adjectives	Nouns	Verbs	All
RNN-80	**10.1**	8.1	**30.4**	**19.0**
CW-50	1.1	2.4	8.1	4.5
CW-100	1.3	4.1	8.6	5.0
HLBL-50	4.4	5.4	23.1	13.0
HLBL-100	7.6	**13.2**	30.2	18.7

Table 3: Comparison of RNN vectors with Turian's Collobert and Weston based vectors and the Hierarchical Log-Bilinear model of Mnih and Hinton. Percent correct.

Results

Semantic Regularities

Method	Spearman's ρ	MaxDiff Acc.
LSA-640	0.149	0.364
RNN-80	0.211	0.389
RNN-320	0.259	0.408
RNN-640	0.270	0.416
RNN-1600	**0.275**	**0.418**
CW-50	0.159	0.363
CW-100	0.154	0.363
HLBL-50	0.149	0.363
HLBL-100	0.146	0.362
UTD-NB	0.230	0.395

Table 4: Results in measuring relation similarity

Results

Semantic Regularities –

- Current state of the art results on SemEval 2 to contextualise the RNN's performance scores.

http://aclweb.org/aclwiki/index.php?title=SemEval-2012_Task_2_%28State_of_the_art%29

Algorithm	Reference	MaxDiff	Spearman
BUAP	Tovar et al. (2012)	31.7	0.014
Random baseline	Jurgens et al. (2012)	31.2	0.018
Duluth-V2	Pedersen (2012)	31.1	0.038
Duluth-V1	Pedersen (2012)	31.5	0.039
Duluth-V0	Pedersen (2012)	32.4	0.050
PMI baseline	Jurgens et al. (2012)	33.9	0.112
UTD-SVM	Rink & Harabagiu (2012)	34.7	0.116
UTD-NB	Rink & Harabagiu (2012)	39.4	0.229
RNN-1600	Mikolov et al. (2013)	41.8	0.275
UTD-LDA	Rink & Harabagiu (2013)	----	0.334
PairDirection	Levy & Goldberg (2014)	45.2	----
Com	Zhila et al. (2013)	45.2	0.353
SuperSim	Turney (2013)	47.2	0.408

Conclusion

- Word representations learnt by RNN language models generally perform well at capturing semantic and syntactic regularities.
- Generally applicable Vector Offset Method for identifying linguistic regularities in continuous space word representations.
- "Byproduct of an unsupervised maximum likelihood training criterion on a large amount of text data."

- Further study:
  - Further testing of the system's robustness
    - Vectors can represent different linguistic regularities – what about testing performance on combinations such as person and number in Verbs?