Digital Theory Lab, NYU
New York, NY, USA

Why Can Computers Understand Natural Language?

The Structuralist Image of Language Behind Word Embeddings

Juan Luis Gastaldi

ETH zürich

February 3rd, 2023

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730

Outline

DNNs and Language

Word Embeddings

How Does It Work?

Why Does It Work?

The Structuralist Background

Outline

DNNs and Language

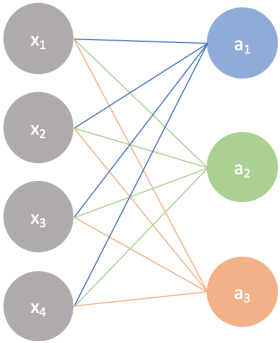
Word Embeddings

How Does It Work?

Why Does It Work?

The Structuralist Background

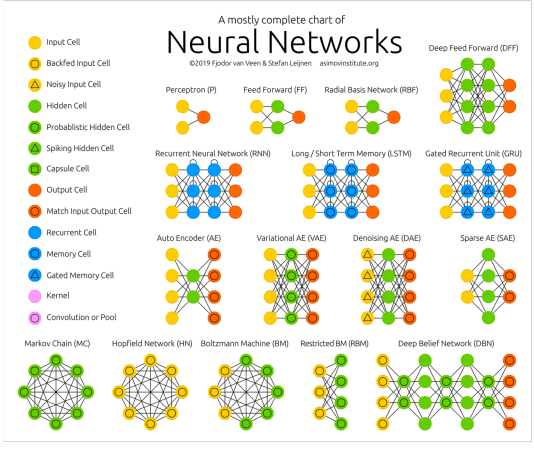
Neural Networks

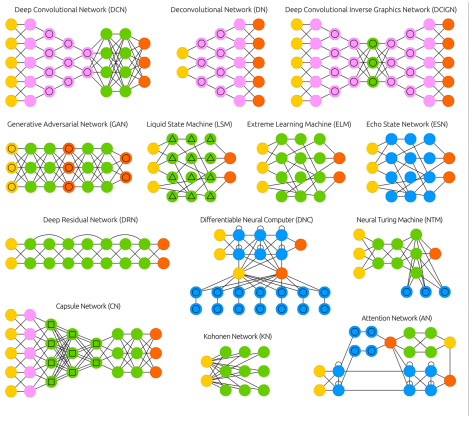


$$\begin{bmatrix} w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \\ w_1 & w_2 & w_3 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} b \\ b \\ b \end{bmatrix} = \begin{bmatrix} w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \\ w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \end{bmatrix} \xrightarrow[activation]{} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$$

Credit: Jeremy Jordan

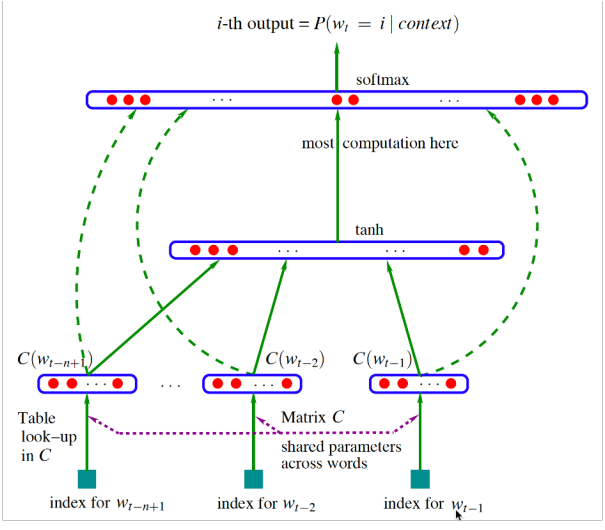
Deep Neural Nets (DNNs)





Source: https://www.asimovinstitute.org/neural-network-zoo/

DNNs and Natural Language



(Bengio et al., 2003)

Outline

DNNs and Language

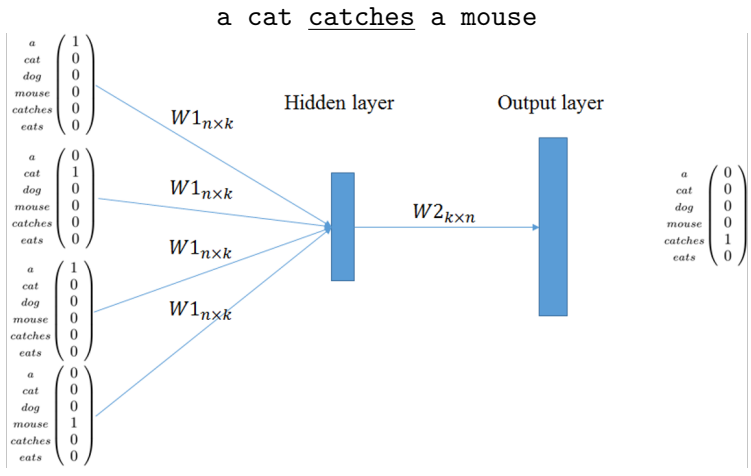
Word Embeddings

How Does It Work?

Why Does It Work?

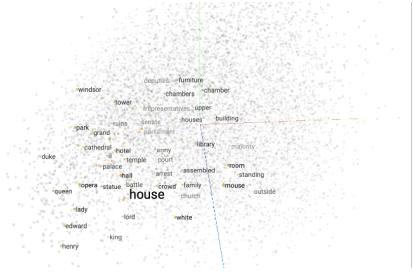
The Structuralist Background

Word Embeddings: word2vec

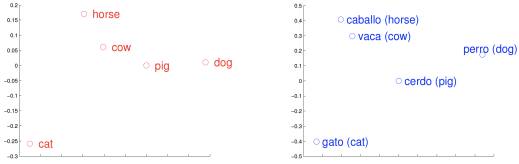


(Source: Ferrone et al., 2017)

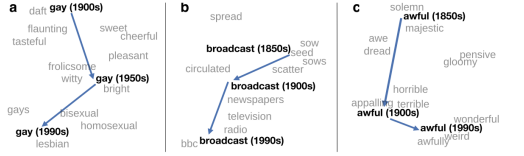
Word Embeddings



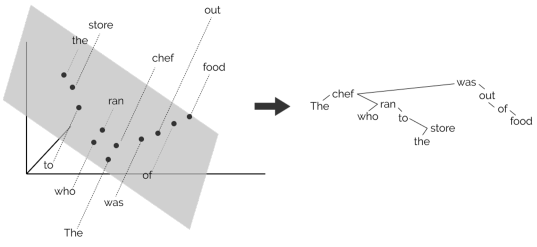
(https://projector.tensorflow.org)



(Mikolov, Sutskever, et al., 2013)



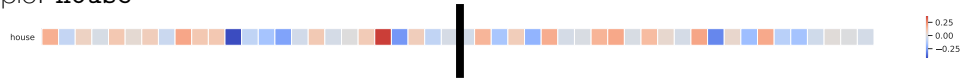
(Hamilton et al., 2016)



(https://nlp.stanford.edu/~johnhew/structural-probe.html)

Word Embeddings: Example

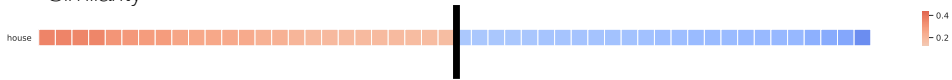
- Example: **`house`**



Word Embeddings: Example

- Example: **`house`**

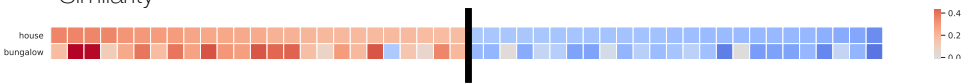
- Syntactic and semantic properties
  - Similarity



Word Embeddings: Example

- Example: **`house`**

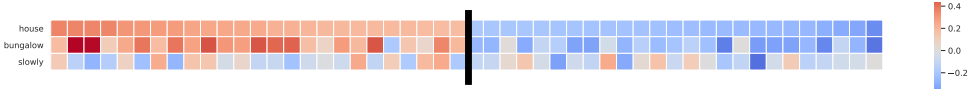
- Syntactic and semantic properties
  - Similarity



Word Embeddings: Example

- Example: **`house`**

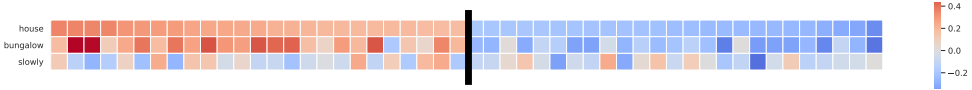
- Syntactic and semantic properties
  - Similarity



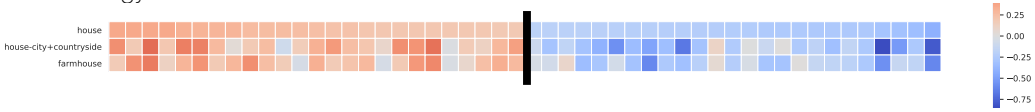
Word Embeddings: Example

- Example: **`house`**

- Syntactic and semantic properties
  - Similarity



  - Analogy

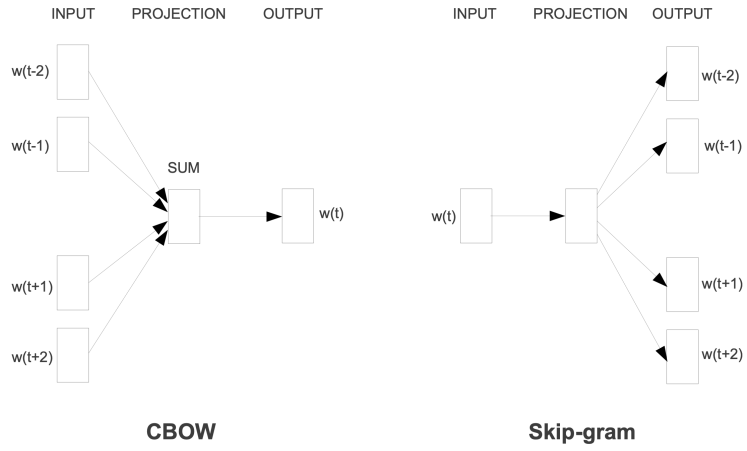


Theoretical Consequences

- The automatic reconstruction of the underlying organization of language does not require more human intervention than the one implied in the most ordinary use of language as recorded in a practically raw linguistic corpus.
- In that reconstruction, both the semantic and syntactic contents of words are determined at once and as the result of the same procedure.
- Word vector representations are not simply disposed in similarity neighbourhoods, but that the vector space itself is also structured following precise directions at the crossroads of which syntactic and semantic contents are established.

Outline

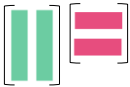
word2vec as Matrix Factorization



(Mikolov, Chen, et al., 2013)

Matrix Factorization

$$A = CR$$



Independent columns in $\boldsymbol{C}$
Row echelon form in $\boldsymbol{R}$
Leads to column rank = row rank

$$A = LU$$

$\boldsymbol{LU}$ decomposition from
Gaussian elimination
(Lower triangular) (Upper triangular)

$$A = QR$$

$\boldsymbol{QR}$ decomposition as
Gram-Schmidt orthogonalization
Orthogonal $\boldsymbol{Q}$ and triangular $\boldsymbol{R}$

$$S = Q\Lambda Q^{\mathrm{T}}$$

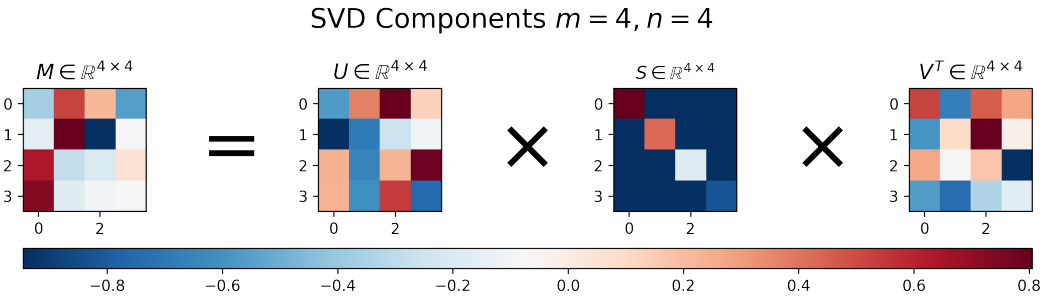
Eigenvalue decomposition
of a symmetric matrix $\boldsymbol{S}$
Eigenvectors in $\boldsymbol{Q}$ eigenvalues in Λ

$$A = U\Sigma V^{\mathrm{T}}$$

Singular value decomposition
of all matrices $\boldsymbol{A}$
Singular values in Σ

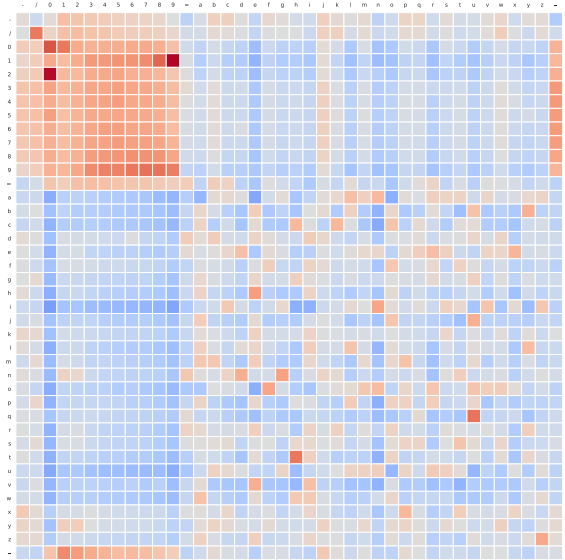
(Hiranabe, 2022)

Singular Value Decomposition (SVD)

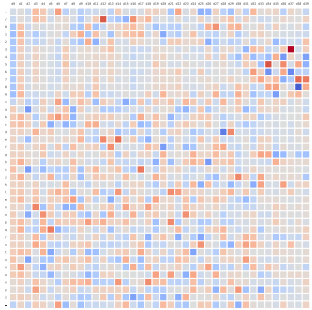


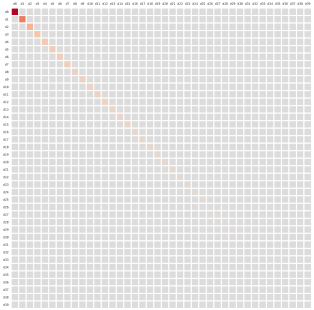
https://dustinstansbury.github.io/theclevermachine/singular-value-decomposition

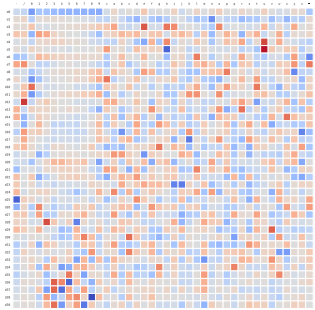
Example: Characters in Wikipedia



Example: Characters in Wikipedia

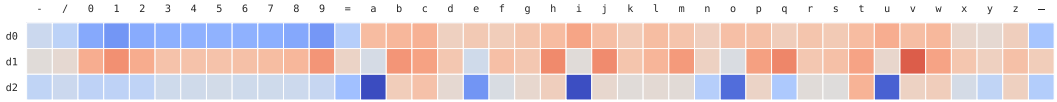




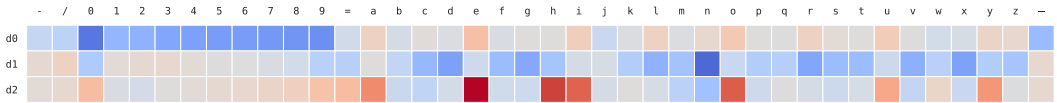


Example: Characters in Wikipedia

Left Singular Vectors:



Right Singular Vectors:



Singular Values:

Outline

DNNs and Language

Word Embeddings

How Does It Work?

Why Does It Work?

The Structuralist Background

Why Can Computers Understand Natural Language?

Faced with the question “why can computers understand natural language?” we should direct our attention to natural language rather than to computers:

“What must natural language be for the specific algebraic procedures of underlying word embeddings to succeed in revealing some of its most essential aspects?”

The Distributional Hypothesis

- “You shall know a word by the company it keeps!” (Firth, 1957)
- “Words which are similar in meaning occur in similar contexts” (Rubenstein & Goodenough 1965)
- “Words with similar meanings will occur with similar neighbors if enough text material is available” (Schütze & Pedersen 1995)
- “A representation that captures much of how words are used in natural context will capture much of what we mean by meaning” (Landauer & Dumais 1997)
- “Words that occur in the same contexts tend to have similar meanings” (Pantel 2005)
- “The degree of semantic similarity between two linguistic expressions A and B is a function of the similarity of the linguistic contexts in which A and B can appear” (Lenci, 2010)

Context and Use

- Theory of (linguistic) meaning as “usage” (Wittgenstein) “the meaning of a word is defined by the circumstances of its use” (Manning and Schütze, 1999)
- Two versions of the Distributional Hypothesis:
  - Weak: Correlation between context and word meaning (Spence and Owens, 1990)
  - Strong: Causality attributed to contextual distributions (Miller and Charles, 1991)
- Context: the domain or scope within which entities of the same nature can be presented together (“co-occur”), in such a way that they can be associated by a cognitive agent.

From Distribution to Structure

a = your c = my	w = apartment x = house y = chair z = stool

your : house
my : apartment

	…	w	x	y	z	…
…	…	0	0	0	0	…
a	…	0	1	1	0	…
b	…	0	0	1	1	…
c	…	1	0	0	1	…
…	…	0	0	0	0	…

Theoretical Consequences

- Contexts are formal dimensions, not empirical circumstances
- Linguistic units are governed by bi-dual relations at different levels
- Language is a game requiring a strategy rather than a tool determined by a use
- Language is a collective phenomenon, expressing organizing principles of a culture

Scientific Orientations

- Semantics and syntax should be studied as a continuum
- The privilege of words should be challenged, in favor of units of all levels (segmentation)
- The study of semantic similarity should lead to the derivation of underlying structures

Outline

DNNs and Language

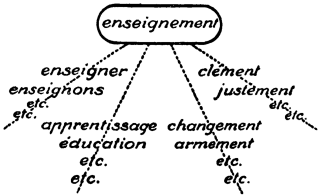
Word Embeddings

How Does It Work?

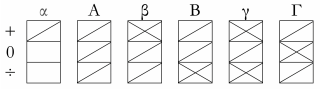
Why Does It Work?

The Structuralist Background

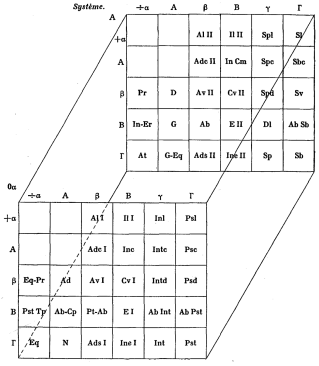
Structuralist Concepts and Tools



(Saussure, 1980)



(Hjelmslev, 1975)



(Hjelmslev, 1935)

Segments	Environments												
	#— r	#— r	#— l	e i — C	æ — C	a o u — C	s — e i	s — æ	s — a o u	…	l —	C[3] —	
l	√												
t		√		√	√	√	√	√	√				
k						√			√				
k		√	√		√			√					
k				√			√						
G						√							
g		√	√		√								
G				√									
r				√	√	√							√
r												√	

(Harris, 1960)

Structuralist Concepts and Tools

Segments	Environments											
	#— r	#—r	#—l	e i —C	æ—C	a o u —C	s— e i	s—æ	s— a o u	…	t —	C³—
t	√											
t		√		√	√	√	√	√	√			
K						√			√			
k		√	√		√			√				
K				√			√					
G						√						
g		√	√		√							
G				√								
r				√	√	√						√
r											√	

(Harris, 1960)

The Structuralist Hypothesis

- Meaning is the effect of structure
- Distributional properties convey meaning only through the action of a latent structure determining possible semantic values, and which is inseparable from the principles of identification of the elementary units of language, since meaning is the effect of discriminating operations performed through segmentation procedures of which the units of language keep the trace
- Linguistic content is the effect of a virtual structure of classes and dependencies at multiple levels underlying (and derivable from) the mass of things said or written in a given language

The Structuralist Hypothesis

We need to recognize a major stake in the possibility […] of understanding language and meaning independently of purely individual practices, and yet not resorting to absolute principles valid without restriction that would make the analysis of actual practices inconsequential. At equal distance from those two positions, [the structuralist image of language] allows us to envisage language as a collective playground, as a reservoir where the significant distinctions resulting from a collective construction of signs are deposited as the most intimate treasure of a culture—to borrow a Saussurean metaphor. It is, after all, the image of those cultures that models like word embeddings give us the means to depict.

Reference Paper

- J. L. Gastaldi. Why Can Computers Understand Natural Language?
  In: *Philosophy & Technology* 34.1 (2021), pp. 149–214.

Related Papers

- J. L. Gastaldi and L. Pellissier. The calculus of language: explicit representation of emergent linguistic structure through type-theoretical paradigms
  In: *Interdisciplinary Science Reviews* 46.4 (2021), pp. 569–590.
- J. L. Gastaldi, Content from Expressions: The Place of Textuality in Deep Learning Approaches to Mathematics
  Under review at *Synthese. SI: Linguistically Informed Philosophy of Mathematics.* Fisseni, B., Kant, D., Sarikaya, D. and Schröder, B. (Eds.).

References I

Bengio, Y., Ducharme, R., Vincent, P., & Janvin, C. (2003). A neural probabilistic language model. *J. Mach. Learn. Res.*, *3*, 1137–1155.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *CoRR*, *abs/1605.09096*.

Harris, Z. (1960). *Structural linguistics*. University of Chicago Press.

Hjelmslev, L. (1935). *La catégorie des cas*. Wilhelm Fink Verlag.

Hjelmslev, L. (1975). *Résumé of a Theory of Language*. Nordisk Sprog-og Kulturforlag.

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *CoRR*, *abs/1301.3781*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., Dean, J., Le, Q., & Strohmann, T. (2013). *Learning representations of text using neural networks. NIPS deep learning workshop 2013 slides*.

Miller, G. A., & Charles, W. G. (1991). Contextual correlates of semantic similarity. *Language and Cognitive Processes*, *6*(1), 1–28. https://doi.org/10.1080/01690969108406936

Saussure. (1980). *Cours de linguistique générale*. Payot.

Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*(5), 317–330. https://doi.org/10.1007/BF01074363

Digital Theory Lab, NYU
New York, NY, USA

Why Can Computers Understand Natural Language?

The Structuralist Image of Language Behind Word Embeddings

Juan Luis Gastaldi

ETH zürich

February 3rd, 2023

This project has received funding from the
European Union's Horizon 2020 research and innovation programme
under grant agreement No 839730