Procesamiento de Lenguaje Natural

Embeddings

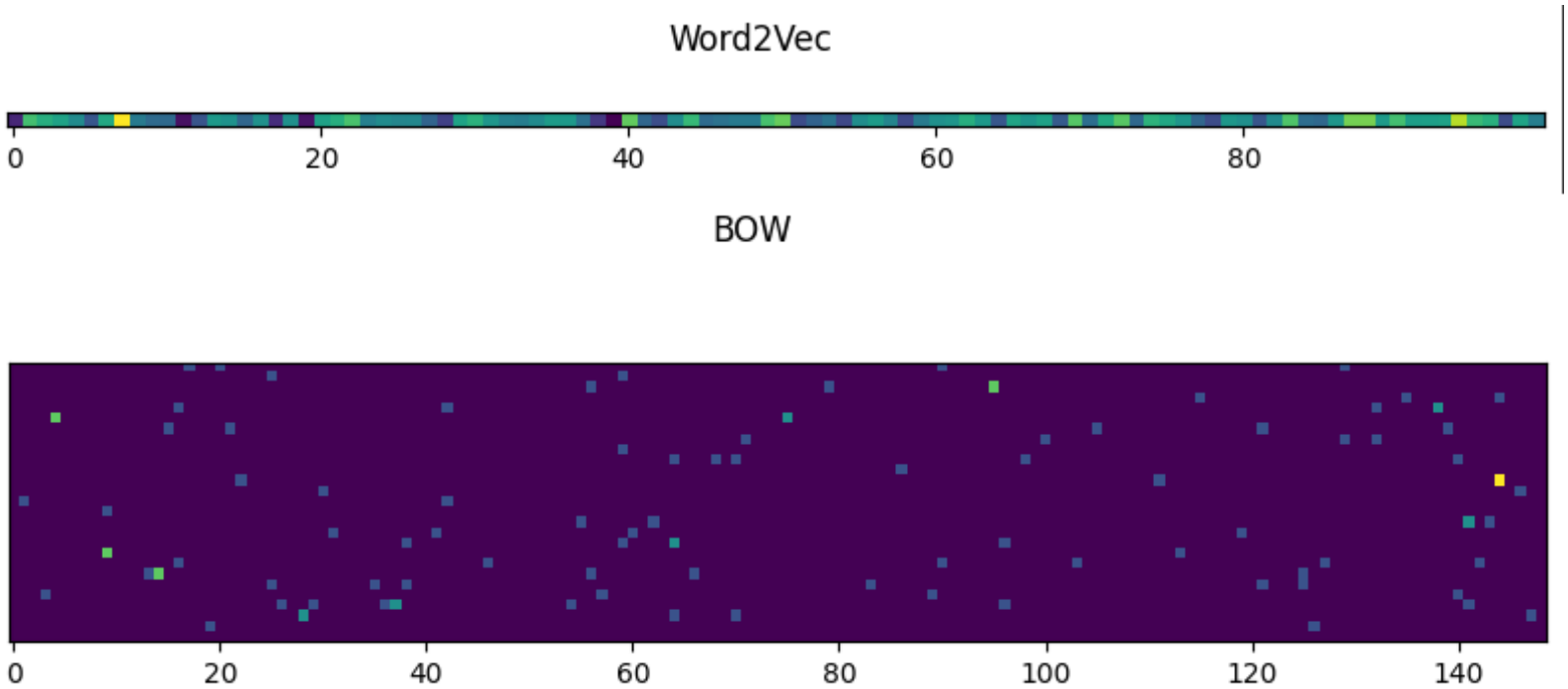
Mauricio Toledo-Acosta
mauricio.toledo@unison.mx

Departamento de Matemáticas
Universidad de Sonora

Section 1

Introducción

Sparsity problem



Bengio, 2003

Journal of Machine Learning Research 3 (2003) 1137–1155 Submitted 4/02; Published 2/03

A Neural Probabilistic Language Model

Yoshua Bengio BENGIOY@IRO.UMONTREAL.CA
Réjean Ducharme DUCHARME@IRO.UMONTREAL.CA
Pascal Vincent VINCENTP@IRO.UMONTREAL.CA
Christian Jauvin JAUVINC@IRO.UMONTREAL.CA
Département d'Informatique et Recherche Opérationnelle
Centre de Recherche Mathématiques
Université de Montréal, Montréal, Québec, Canada

Editors: Jaz Kandola, Thomas Hofmann, Tomaso Poggio and John Shawe-Taylor

The original paper

Antecedentes, 1991

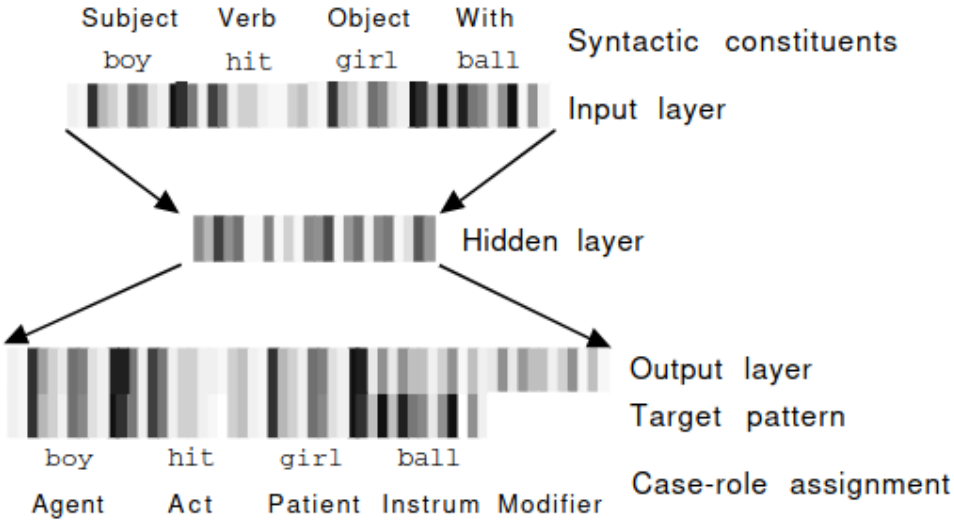


Figure 2: **Snapshot of basic FGREP simulation.** The input and output layers of the network are divided into assemblies, each holding one word representation at a time. Each unit in an input assembly is set to the activity value of the corresponding component in the lexicon entry. The input layer is fully connected to the hidden layer and the hidden layer to the output layer. Connection weights are omitted from the figure. If the network has successfully learned the task, each output assembly forms an activity pattern identical to the lexicon representation of the word filling that role. The correct role assignment is shown at the bottom of the display. This pattern forms the output target for the network. Grey-scale values from white to black are used in the figure to code the unit activities, which vary within the range [0,1].

Antecedentes, 1991

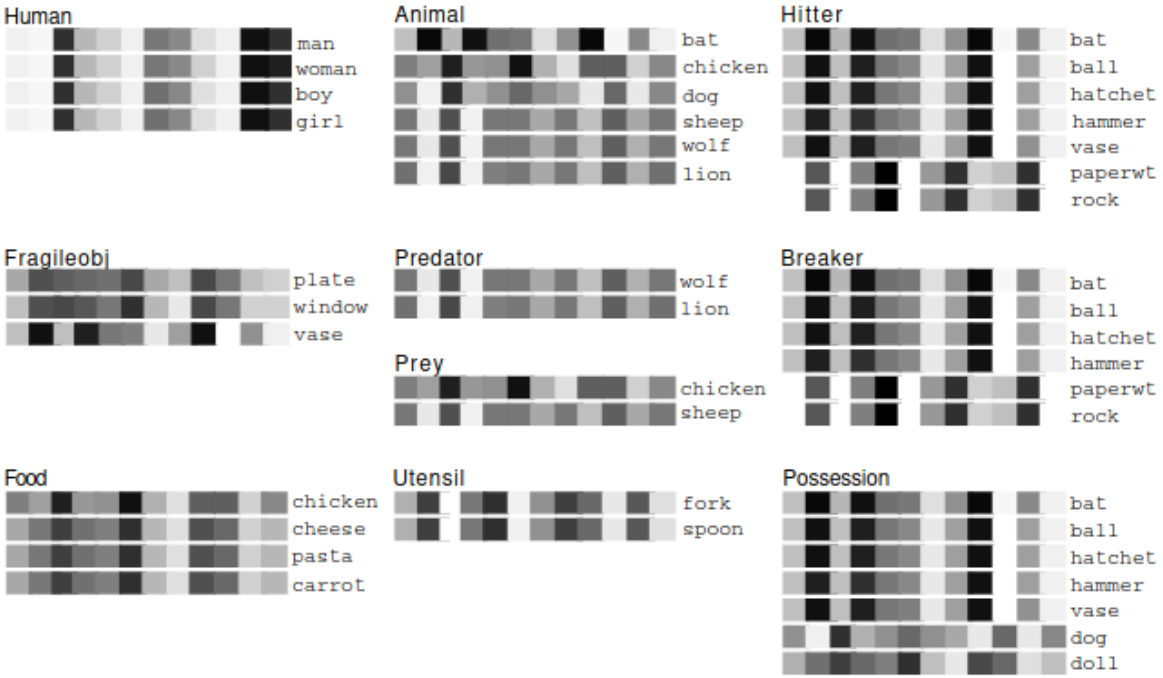


Figure 3: **Final representations.** The representations for the synonymous words {`man`, `woman`, `boy`, `girl`}, {`fork`, `spoon`}, {`wolf`, `lion`}, {`plate`, `window`}, {`ball`, `hatchet`, `hammer`}, {`paperwt`, `rock`} and {`cheese`, `pasta`, `carrot`} have become almost identical.

Word2Vec

Sentence	**He poured himself a cup of coffee**
Target	**himself**

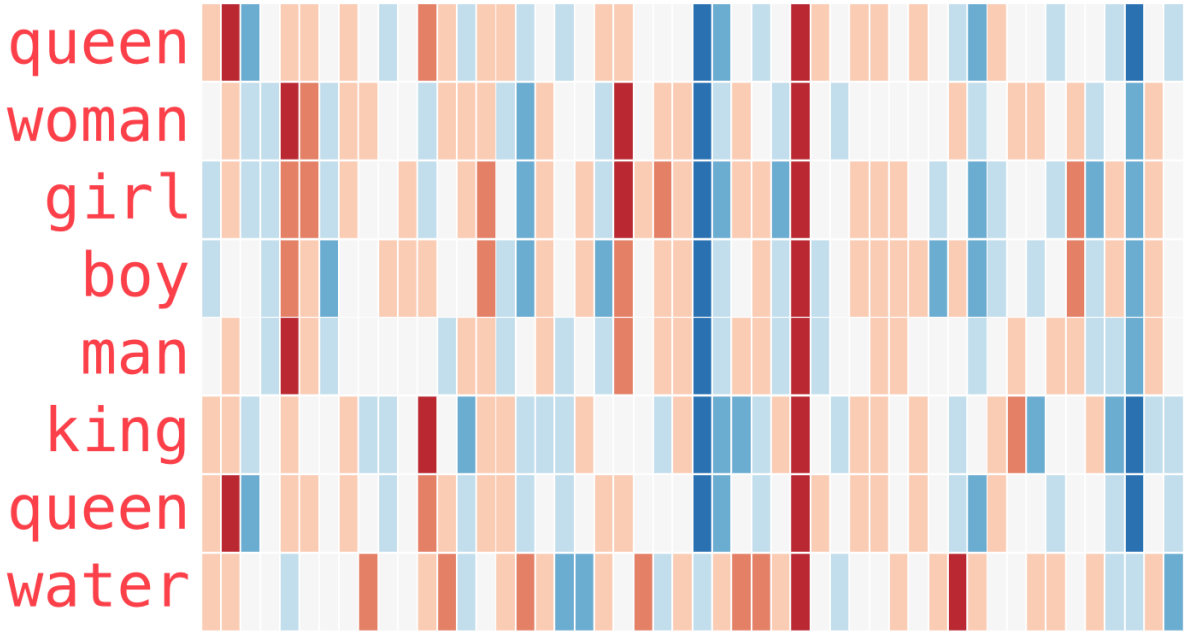
- Continuous Bag-Of-Words

  input *He, poured, a, cup*
  output *himself*

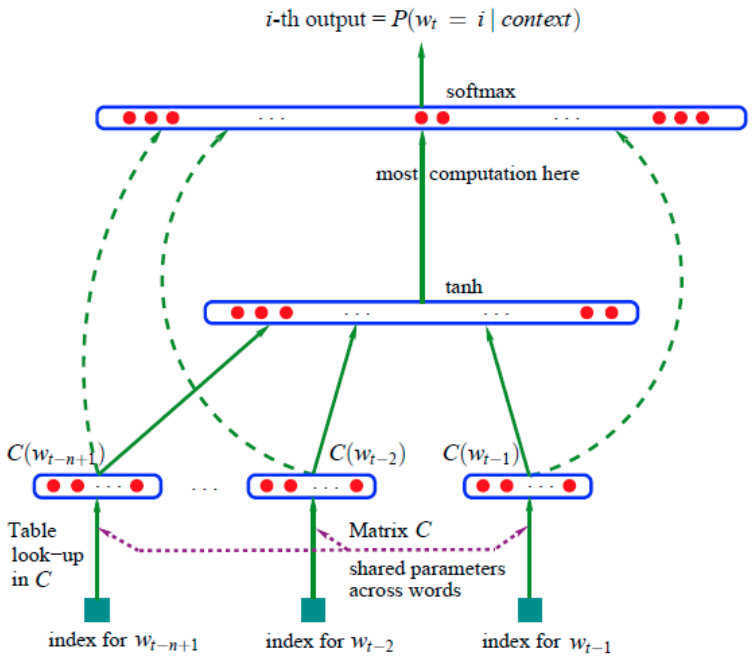
- Skip-gram model

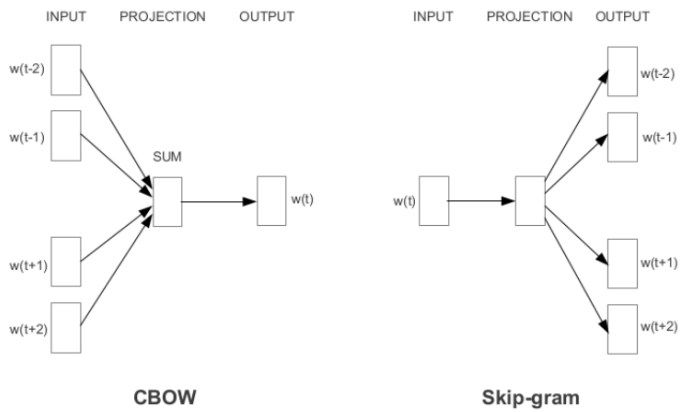
  input *himself*
  output *He, poured, a, cup*

Original Paper


queen
woman
girl
boy
man
king
queen
water

Diferencias entre los enfoques de NPLM y Word2Vec

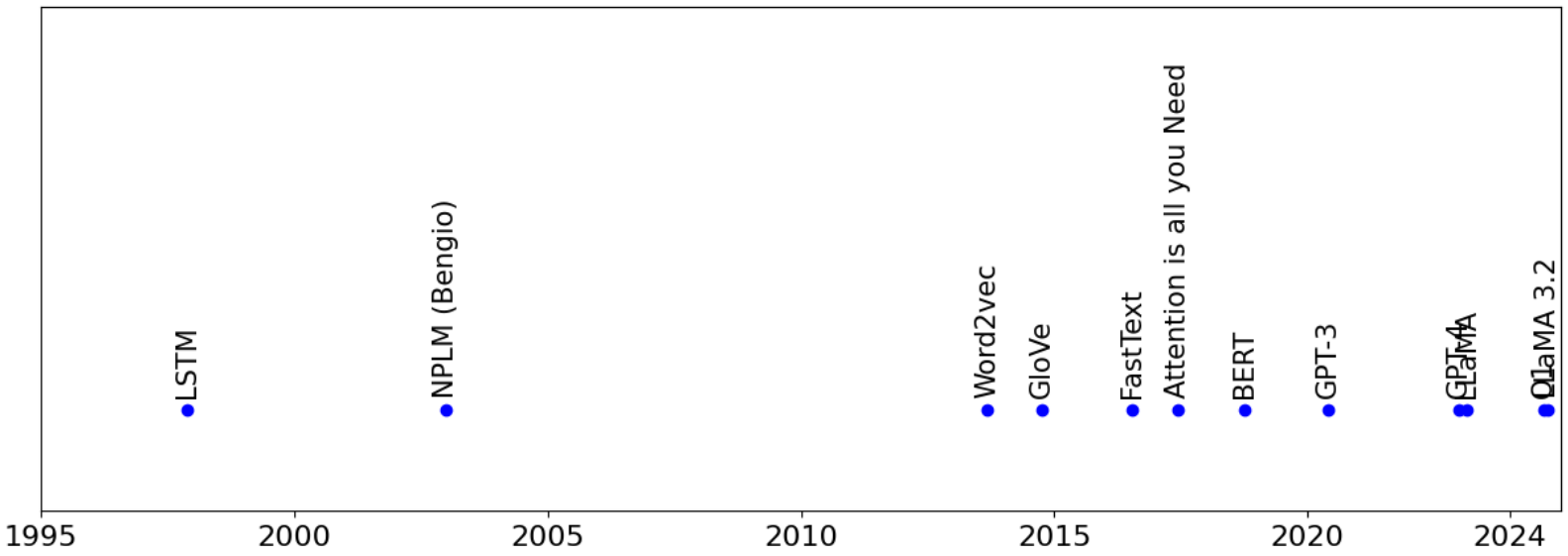




GloVe

In contrast to word2vec, GloVe seeks to make explicit what word2vec does implicitly: Encoding meaning as vector offsets in an embedding space – seemingly only a serendipitous by-product of word2vec – is the specified goal of GloVe.

Timeline


LSTM
NPLM (Bengio)
Word2vec
GloVe
FastText
Attention is all you Need
BERT
GPT-3
1995
2000
2005
2010
2015
2020
2024