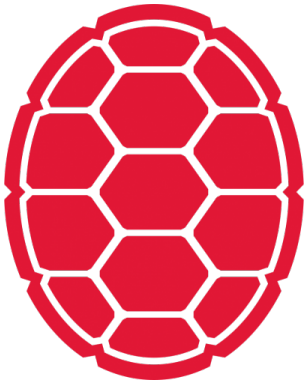


Representation Learning

Machine Learning: Jordan Boyd-Graber
University of Maryland
SLIDES ADAPTED FROM YOAV GOLDBERG AND OMER LEVY

Word Representation

- Last time, we saw how valuable hidden layers were for representation
- How can we use it for words, images, etc.?

Word Representation

- Last time, we saw how valuable hidden layers were for representation
- How can we use it for words, images, etc.?
- How similar is “pasta” to “pizza”
- Computers often use one-hot representations
- Or fragile knowledge bases

Word Representation

- Last time, we saw how valuable hidden layers were for representation
- How can we use it for words, images, etc.?
- How similar is “pasta” to “pizza”
- Computers often use one-hot representations
- Or fragile knowledge bases
- Distributional Hypothesis (Harris, 1954; Firth, 1957)
- Know the word by the company it keeps

Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

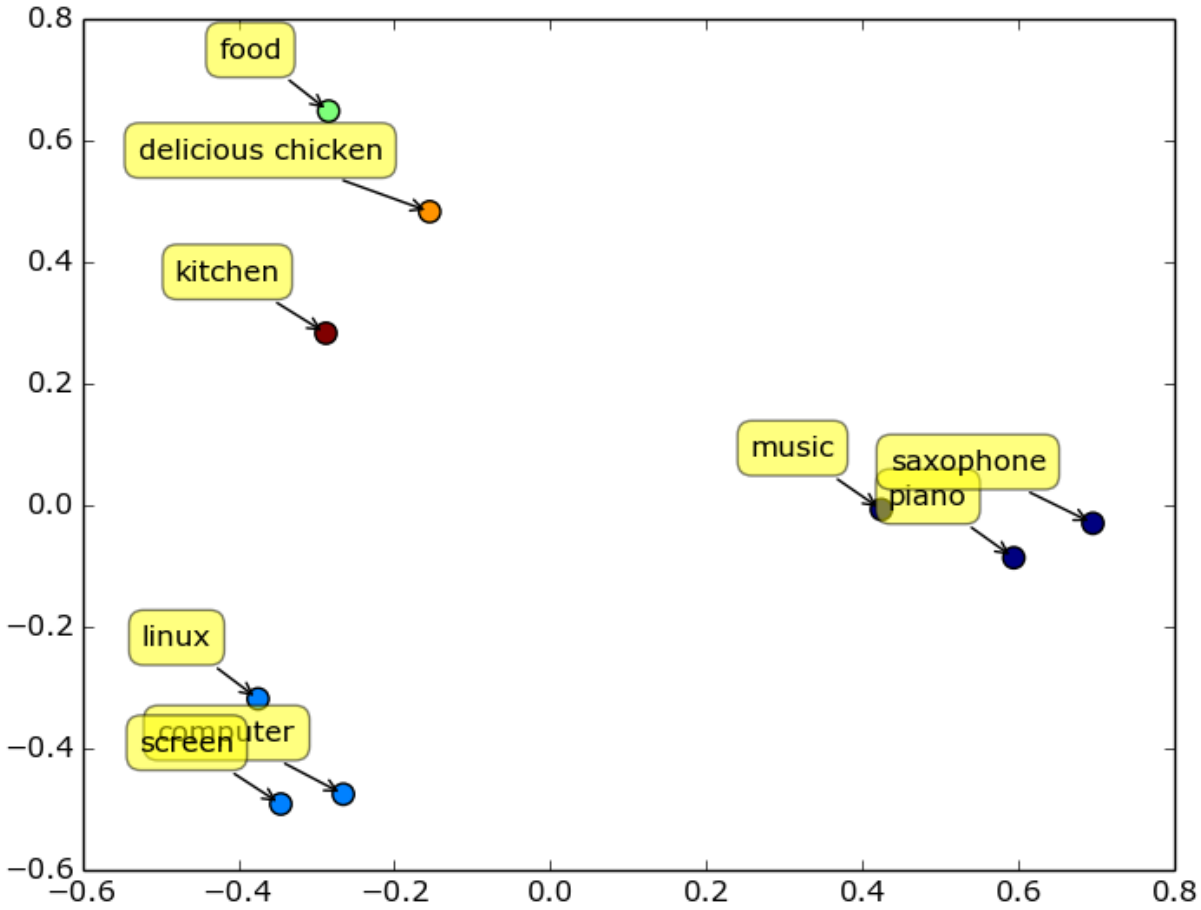
Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

Representation



Usefulness

- Multimodal
- Multilingual
- Useful downstream feature

From Distributional to Distributed Semantics

The new kid on the block

- Deep learning / neural networks
- "Distributed" word representations
  - Feed text into neural-net. Get back "word embeddings".
  - Each word is represented as a low-dimensional vector.
  - Vectors capture "semantics"
- `word2vec` (Mikolov et al)