

Distributional Semantics

Jordan Boyd-Graber

University of Maryland

Slides Adapted from Yoav Goldberg and Omer Levy

Word Representation

- Before, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?

Word Representation

- Before, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?
- How similar is “pasta” to “pizza”
- Computers often use one-hot representations
- Or fragile knowledge bases

Word Representation

- Before, we saw how valuable hidden layers were for representation (much more language today)
- How can we use it for words?
- 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

Obvious things

- Use images?

Obvious things

- Use images?
 - ▶ How our eyes do it!
 - ▶ We lose information
 - ▶ OCR is often preprocessing step
- Use strings?

Obvious things

- Use images?
 - ▶ How our eyes do it!
 - ▶ We lose information
 - ▶ OCR is often preprocessing step
- Use strings?
 - ▶ Wasteful of memory
 - ▶ Is dOg different from Dog?
 - ▶ What about “ dog” and “dog”?

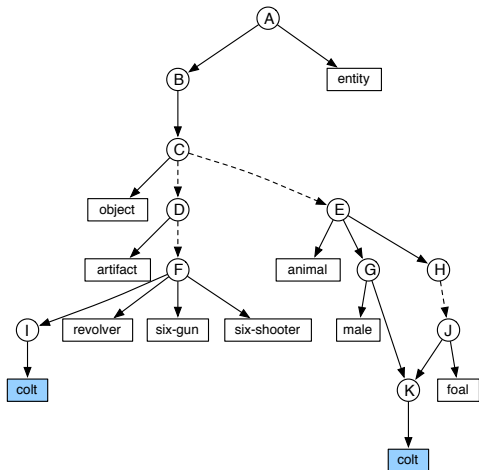
What we've already done

- Typically, want to do preprocessing (case, whitespace)
- Can also remove plurals, verb forms, etc. (more later)
- Then, you can represent each word as an **integer**

What we've already done

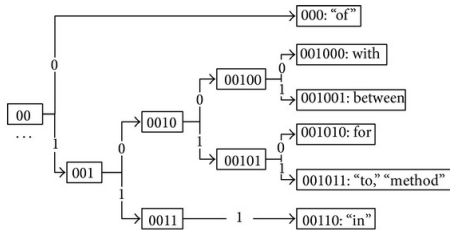
- Typically, want to do preprocessing (case, whitespace)
- Can also remove plurals, verb forms, etc. (more later)
- Then, you can represent each word as an **integer**
 - ▶ Memory efficient
 - ▶ Removes connections between words

What about a dictionary?



- **WordNet:**
electronic dictionary
- Brown clusters:
automatically built tree
- Lesk algorithm:
from dictionary (use tf-idf cosine)

What about a dictionary?



- WordNet: electronic dictionary
- **Brown clusters:** automatically built tree
- Lesk algorithm: from dictionary (use tf-idf cosine)

What about a dictionary?

Pizza

a dish of Italian origin consisting of a flat, round base of dough baked with a topping of tomato sauce and cheese, typically with added meat or vegetables.

Pasta

a dish originally from Italy consisting of dough made from durum wheat and water, extruded or stamped into various shapes and typically cooked in boiling water.

- WordNet: electronic dictionary
- Brown clusters: automatically built tree
- Lesk algorithm: from dictionary (use tf-idf cosine)

What about a dictionary?

Pizza

a **dish** of **Italian** origin consisting of a flat, round base of dough baked with a topping of tomato sauce and cheese, typically with added meat or vegetables.

Pasta

a **dish** originally from **Italy** consisting of dough made from durum wheat and water, extruded or stamped into various shapes and typically cooked in boiling water.

- WordNet: electronic dictionary
- Brown clusters: automatically built tree
- **Lesk algorithm**: from dictionary (use tf-idf cosine)

What about a dictionary?

Pizza

a dish of Italian **origin** consisting of a flat, round base of dough baked with a topping of tomato sauce and cheese, typically with added meat or vegetables.

Pasta

a dish **originally** from Italy consisting of dough made from durum wheat and water, extruded or stamped into various shapes and typically cooked in boiling water.

- WordNet: electronic dictionary
- Brown clusters: automatically built tree
- Lesk algorithm: from dictionary (use tf-idf cosine)
- Infinitely deep rabbit hole!

Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

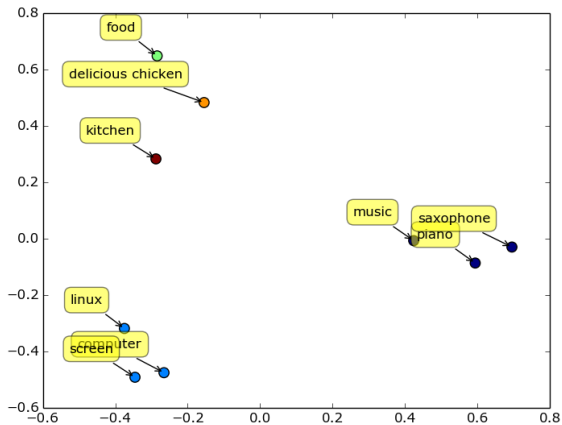
Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

Intuition (from Boroni)

Marco saw a furry little wampimuk hiding in the tree

Continuous Representation!



Definitions

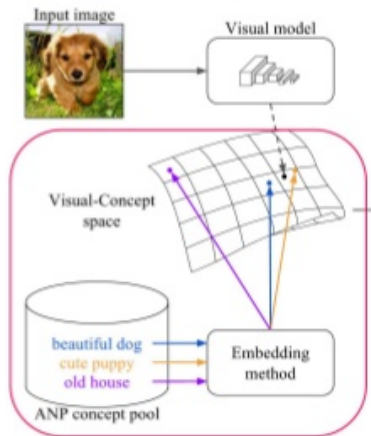
Distributional Word Representation

Encoding a word's meaning from how it is used in a corpus

Distributed Representation

Encoding a word's meaning in a compact, dense, low-dimensional representation

Usefulness



- Multimodal
- Multilingual
- Useful downstream feature