https://qdata.github.io/deep2Read/

Retrofitting Word Vectors to Semantic Lexicons

Manaal Faruqui Jesse Dodge Sujay K. Jauhar Chris Dyer Eduard Hovy Noah A. Smith

Language Technologies Institute, CMU

Presentation by Jack Morris

Word embeddings

- A system that maps words from a vocabulary to meaningful vectors
- Data driven: trained on text
- Popular word embeddings: word2vec, glove
- Can be used for identifying relationships between pairs of words

GloVe Demo

Finding the nearest neighbors of sweden

```
nearest_neighbors
('sweden')
```

Word	Cosine distance
norway	0.760124
denmark	0.715460
finland	0.620022
switzerland	0.588132
belgium	0.585835
netherlands	0.574631
iceland	0.562368
estonia	0.547621
slovenia	0.531408

Semantic lexicons

- Human-written rules about vocabulary words and their relationships to one another
- Contain synonyms, hypernyms, hyponyms, paraphrase relationships
- Popular semantic lexicons: WordNet, FrameNet, Paraphrase Database

WordNet Demo

Finding hypernyms and hyponyms of **paint**

Noun

- S: (n) **paint**, pigment (a substance used as a coating to protect or decorate a surface (especially a mixture of pigment suspended in a liquid); dries to form a hard coating) *"artists use `paint' and `pigment' interchangeably"*
  - *direct hyponym* / *full hyponym*
    - S: (n) acrylic, acrylic paint (used especially by artists)
    - S: (n) antifouling paint (a paint used to protect against the accumulation of barnacles etc. on underwater surfaces)
    - S: (n) coat of paint (a layer of paint covering something else)
    - S: (n) distemper (paint made by mixing the pigments with water and a binder)
    - S: (n) enamel (a paint that dries to a hard glossy finish)
    - S: (n) encaustic (a paint consisting of pigment mixed with melted beeswax; it is fixed with heat after application)
    - S: (n) finger paint, fingerpaint (paint that has the consistency of jelly)
    - S: (n) house paint, housepaint (paint used to cover the exterior woodwork of a house)
    - S: (n) oil paint (paint in which a drying oil is the vehicle)
    - S: (n) semigloss (a paint that dries with a finish between glossy and flat)
    - S: (n) spray paint (paint applied with a spray gun)
    - S: (n) water-base paint (paint in which water is used as the vehicle)
  - *substance meronym*
  - ***direct hypernym*** / *inherited hypernym* / *sister term*
    - S: (n) coating, coat (a thin layer covering something) *"a second coat of paint"*
    - S: (n) coloring material, colouring material, color, colour (any material used for its color) *"she used a different color for the trim"*

Problem: Improving word embeddings

- Unsupervised learning of word embeddings is good, but has problems
- Plus, we have databases of carefully arranged categorized information about our language (semantic lexicons)
- How can we use our domain knowledge to improve word embeddings?

Solution

Retrofitting Word Vectors to Semantic Lexicons

Idea: balance distance from original vector in embedding space with distance between word and its associations from semantic lexicon

Can do this by minimizing:

$$\Psi(Q) = \sum_{i=1}^{n} \left[\alpha_i \|q_i - \hat{q}_i\|^2 + \sum_{(i,j) \in E} \beta_{ij} \|q_i - q_j\|^2 \right]$$

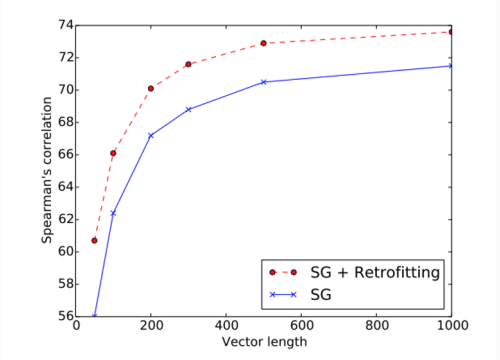
Turns out, ψ is convex in Q, so we can solve it using gradient descent

Retrofitting word vectors: **Results**

Retrofitted vectors **outperform** their original counterparts, as well as prior work that tried to incorporate information from semantic lexicons during training

The process is very **fast** -- takes 5 seconds for a graph of 100,000 words and vector length 300

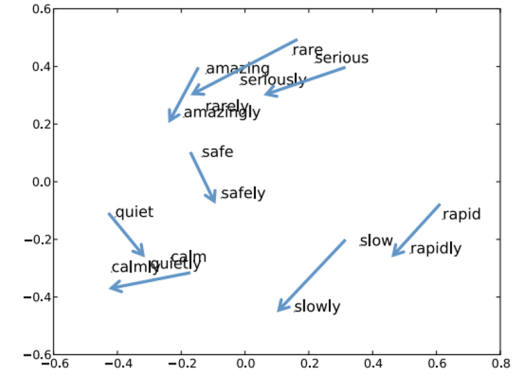
Multilingual: this technique works on embeddings & lexicons from any language



Retrofitting word vectors: **Results**

Modeling relationships: “adjective to adverb”

before



after