Compositional Distributional Semantic Models for Semantic Relatedness and Entailment

Sidharth Gupta
Sai Krishna Prasad

Guided By:- Amitabha Mukherjee

Distributional Semantic Models (DSMs)

- Distributional hypothesis - words that occur in the same context tend to have similar meanings
- Firth - “a word is characterised by the company it keeps”
- Collect distributional information for words in a corpus in high-dimensional vectors
- Unsupervised learning of vectors for words
- Semantic similarity for words - define in terms of vector similarity

Compositional DSMs

- How to combine the meanings of words, to understand the semantics of full sentences?
- Extend DSMs - compositionality
- Simple approaches:
  - Weighted sum of vectors
  - Element wise product of vectors
  - Commutative, no attention to syntax
- Operator words - modify the meanings of other words in their context (adjectives, transitive verbs)
- Model these as matrices - “act” on the meanings of other words

DataSet

- SICK Database
- 10,000 english sentence pairs divided equally between the training and test data sets
- The training data contains the following fields
  1. sentence_A
  2. sentence_B
  3. relatedness_score
  4. entailment_judgment - Entailment, Neutral or Contradiction

Task

- SemEval 2014 - task 1
- SubTask 1 - output the degree of relatedness between two sentences
- SubTask 2 - output the semantic entailment holding between two sentences
- Relatedness score in the training data - average score given by 10 human beings collected for each pair.
- Entailment label - majority label of 5 human beings

Categorical CDSMs

- Grefenstette and Sadrzadeh, 2011[2]
- Pregroup grammars specify syntax for sentences/phrases in the language
- Pregroup grammars - associate types (atomic or compound) with all words in the lexicon
- Eg. cats [n] like [$n^r s n^l$] milk [n]
- Syntax guided semantic composition
- Using distribution information for words provided by a DSM, construct matrices for relational words

Categorical CDSMs

- Matrices for a relational word *P*
  - dimensionality *mr* (*r x r x … r* - *m* times)
  - *m* - adjoint types specified by grammar
  - sum over all instances in corpus appropriate element from the corresponding word vectors ($w_1, w_2, \ldots, w_m$)
- Sentence vector computation
  - Elementwise product over the the matrix for *P* and the appropriate element from $w_1 \times w_2 \times \ldots \times w_m$

Recursive Matrix-Vector Spaces

- A word is represented using a vector and a matrix
- The vector contains the meaning of the word ($a = R^n$)
- The matrix Captures how the word changes the meaning of neighbouring words or phrases.($A = n*n$)
- A composition of two words is represented as
  $p = f(a, b, R, K) = g\left(W \begin{bmatrix} Ba \\ Ab \end{bmatrix}\right)$
  $P = W_M \begin{bmatrix} A \\ B \end{bmatrix}$
  Where R is a known syntactic relation,
  K is background knowledge, and
  W and W_m are (n*2n) matrices

Recursive Matrix-Vector Spaces

- The model generalizes many earlier models such as
  1. Mitchell and Lapata where p = Ba + Ab
  2. Baroni and Zamparelli where p = Ab
  3. Socher (2011) where p = a + b
- $\theta = (W, W_M, W^{label}, L, L_M)$
  Learning is done by using gradient descent method over the parameter space
- To reduce the dimensionality we represent
  A = U*V + dia(a)
- It is also the only model that properly negates the sentiment

References

1. Socher, Richard, Brody Huval, Christopher Manning, and Andrew Ng, 2012. Semantic compositionality through recursive matrix--vector spaces. In Proceedings of EMNLP.
2. Grefenstette, Edward and Mehrnoosh Sadrzadeh, 2011. Experimental support for a categorical compositional distributional model of meaning. In Proceedings of EMNLP.
3. Mitchell, Jeff and Mirella Lapata, 2008. Vector -based models of semantic composition. In Proceedings of ACL. Columbus, OH.
4. Mitchell, Jeff and Mirella Lapata, 2010. Composition in distributional models of semantics. Cognitive Science, 34(8): 1388–1429.

Examples

- A man is jumping into an empty pool.
  There is no biker jumping in the air.
  Score :- 1.6
- A person in black jacket is doing tricks on the motorbike.
  A man in black jacket is doing tricks on the motorbike.
  Score :- 4.9
- Two teams are competing in a football match.
  Two groups of people are competing in a football match.
  Entailment
- The brown horse is near a red barrel.
  The brown horse is far from a red barrel.
  Contradiction
- A man in black jacket is doing tricks on a motorbike.
  A man is riding a cycle.
  Neutral