Mixture of Topic-based Distributional Semantic and Affective Models

Fenia Christopoulou, Eleftheria Briakou,
Elias Iosif, Alexandros Potamianos

School of Electrical and Computer Engineering
National Technical University of Athens, Greece

February 2, 2018

Overview

Goal - Motivation

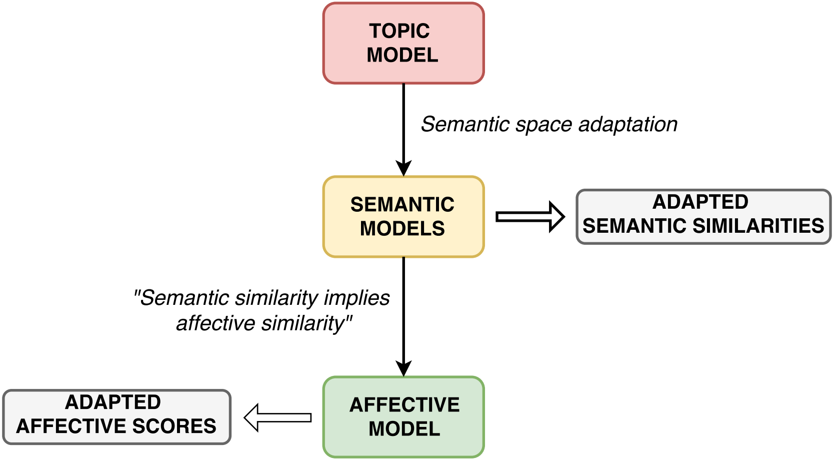
- Goal: Tackle **word sense ambiguity** in
  1. Word-level semantic similarity computation (with/without context information)
  2. Sentence-level affect estimation

- Motivation:
  - Topic domain of sentences influences the meaning of words
  - Limitations of traditional DSMs: One semantic representation → flattened senses

- Prior Work:
  - Sense-agnostic representations [Reisinger and Mooney, 2010]
  - Extended SkipGram word2vec [Neelakantan et al., 2014]
  - Mixture models [Xiang et al., 2014]
  - Latent Dirichlet Allocation (LDA) [Liu et al., 2015]
  - Knowledge bases [Liu et al., 2015; Pilehvar and Collier, 2016]
  - Joint NN Learning [Lin and He, 2009; Zheng et al., 2017]

Overview

Proposed Approach - Overview



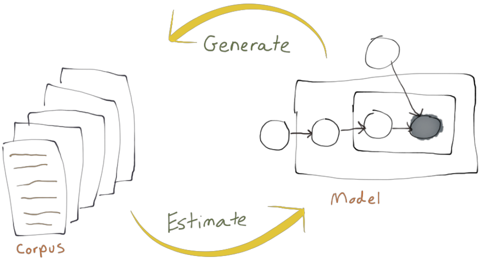
Model Overview

1. Train a probabilistic topic model (LDA)
   - Generic domain documents (corpus)
2. Apply trained model to the same corpus
   - Sentence-wise (**assumption**: one sentence contains one topic)
3. Classify corpus sentences into Topic-specific subcorpora
   - Topic-based posterior probabilities thresholding
4. Train topic-specific DSMs on subcorpora
5. Estimate pairwise word similarities
   - Mixture of semantic word similarities

Topic Modeling

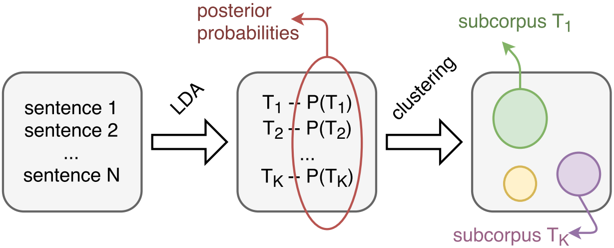
Latent Dirichlet Allocation (LDA) algorithm [Blei et al., 2003]:

- Generative process
- Topic resembles thematic domain
- Document collection as a probabilistic mixture of topics
- Topic as a distribution over words in the collection

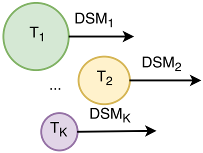


Semantic Space Adaptation

1. Construct sub-corpora using probability-based threshold



2. Train multiple topic-specific DSMs
   - Different topic-based semantic spaces



Overview

Context-independent metrics

Assumption: All topics contribute equally

$$S_{\text{AvgSim}}(w_i, w_j; L_T) = \frac{1}{T} \sum_{t=1}^{|T|} S_t(w_i, w_j; \lambda_t)$$

$$S_{\text{MaxSim}}(w_i, w_j; L_T) = \max_{t \in T} \{S_t(w_i, w_j; \lambda_t)\}$$

where

- L_T set of T topic-specific DSMs
- $S_t(w_i, w_j; \lambda_t)$ semantic similarity of w_i and w_j from λ_t DSM

Context-dependent metrics

Assumption: Topics are weighted with posteriors when context is present

$$S_{\text{AvgSimC}}(w_i, w_j; L_T) = \frac{\sum_{t=1}^{|K(c)|} p(t|c)\, S_t(w_i, w_j; \lambda_t)}{\sum_{t=1}^{|K(c)|} p(t|c)}$$

$$S_{\text{MaxSimC}}(w_i, w_j; L_T) = S_{\hat{t}}(w_i, w_j; \lambda_{\hat{t}})$$

$$\hat{t} = \operatorname*{argmax}_{t \in K(c)} \{p(t|c)\}$$

where

- L_T: set of T topic-specific DSMs
- $S_t(w_i, w_j; \lambda_t)$: semantic similarity of w_i and w_j from λ_t DSM
- $c = c(w_i) \oplus c(w_j)$: shared context of word pair
- $p(t|c)$: posterior probability of topic t for context c
- $K(c) \leq T$: candidate topics with posterior probability > 0.01

Fusion of Topic Models

Motivation:

- Combine information from multiple topic models trained on different number of topics
- Actual number of word senses can be better approached

Assumption: Document collection contains multiple topic distributions

$$S_{\text{Fuse}}(w_i, w_j) = \max_{L_T \in G} \{ S_{*\text{Sim}}(w_i, w_j; L_T) \}$$

where

- $S_{*\text{Sim}}(w_i, w_j; L_T)$: w_i, w_j pair similarity
- G: group of DSM sets to be fused

Linear combination of topic similarities

Motivation: Learn a linear combination of topic-similarities

Assumption: Document collection contains single topic distribution

Expectation: Better estimation of pairwise similarities compared to un-weighted average

$$S_{\mathrm{LR}}(w_i, w_j; L_T) = \beta_0 + \sum_{t=1}^{|T|} \beta_t \, S_t(w_i, w_j; \lambda_t)$$

where

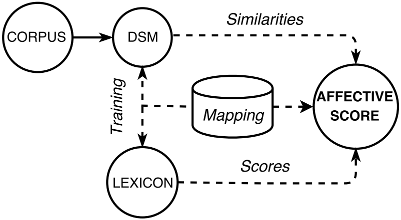
- β_t: learned weight for topic t
- β_0: bias weight
- $S_t(w_i, w_j; \lambda_t)$: similarity of w_i, w_j pair from λ_t DSM

Overview

Semantic-Affective Model

Semantic similarity implies affective similarity

- Affective space: valence
- Affective lexicon: seed words
- DSM: general-purpose corpus
- Semantic-affective mapping [Malandrakis et al., 2011]



$$v(w_j) = \alpha_0 + \sum_{n=1}^{N} \alpha_i \; v(s_i) \; S(s_i, w_j; \lambda)$$

where

- $v(w_j)$: valence score of unknown word w_j
- $v(s_i)$: valence score of seed word s_i
- $S(s_i, w_j; \lambda)$: semantic similarity from λ DSM
- α_i / α_0: weight of seed word s_i/bias weight

Affective Mixture Model

Two-step process:

1. Select topics for each sentence
2. Compute adapted affective scores

$$v_{\text{adapt}}(w_j) = \alpha_0 + \sum_{n=1}^{N} \alpha_i \; v(s_i) \; S_{\text{AvgSimC}}(s_i, w_j; L_T)$$

where

- $v(s_i)$: valence score of seed word s_i
- α_i / α_0: weight of seed word s_i/bias weight
- $S_{\text{AvgSimC}}(s_i, w_j; L_T)$: adapted semantic similarity
- $v_{\text{adapt}}(w_j)$: final adapted valence score for a sentence word

Overview

Data

- Corpora: Web Corpus [Iosif and Potamianos, 2015], Wikipedia[1]
- Affective Lexica: ANEW [Bradley and Lang, 1999]
- Datasets:
  - Word-level Semantic Similarity

Dataset	Pairs	Type
MEN [Bruni et al., 2014]	3000	out-of-context
WS-353 [Finkelstein et al., 2001]	353	out-of-context
SCWS [Huang et al., 2012]	2003	in-context

  - Sentence-level affect estimation

Dataset	Sentences	Valence
SemEval 2007 Task 14 [Strapparava and Mihalcea, 2007]	1000	[-1,1]

[1] https://dumps.wikimedia.org/enwiki/20160720/

Tools & Parameters

- LDA: Gensim Toolbox [Řehůřek and Sojka, 2010]
  - Up to 100 topics
- DSMs: Continuous Bag-of-Words (CBOW) word2vec[2]
  - Corpora: Web corpus, Wikipedia
  - Dimensionality: 300 (Web corpus), 500 (Wikipedia)
  - Context window size: 5
- Semantic similarity metric: cosine
- Evaluation Metric: Spearman's ρ correlation coefficient

[2] https://code.google.com/archive/p/word2vec/

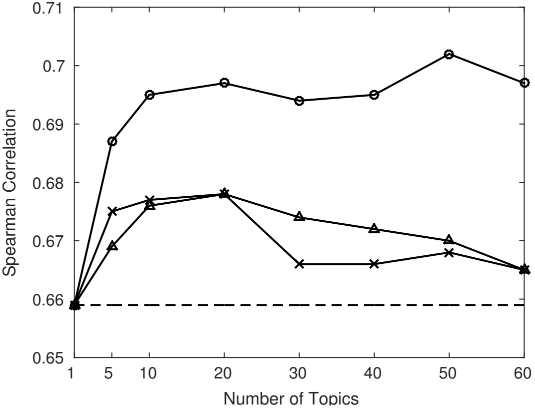
Semantic Similarity Results I

Performance comparison in terms of Spearman's ρ correlation.

	Out-of-Context		*In-Context*		
Approach	WS-353	MEN	SCWS		
			MaxSimC	AvgSim	AvgSimC
Iacobacci et al. [2015]	**0.779**	0.805	0.589	–	0.624
Pilehvar and Collier [2016]	–	0.786	–	**0.708**	**0.715**
Amiri et al. [2016]	–	–	–	–	0.709
Web Corpus					
TDSMs	0.722	0.800	0.678	0.678	0.702
TDSMs-Fuse	–	–	0.674	0.676	0.705
TDSMs-LR	0.727	**0.838**	–	–	–
No Topics	0.703	0.773	0.659		
Wikipedia Corpus					
TDSMs	0.698	0.753	**0.683**	0.696	0.701
TDSMs-Fuse	–	–	0.681	0.685	0.707
TDSMs-LR	0.695	0.796	–	–	–
No Topics	0.644	0.731	0.669		

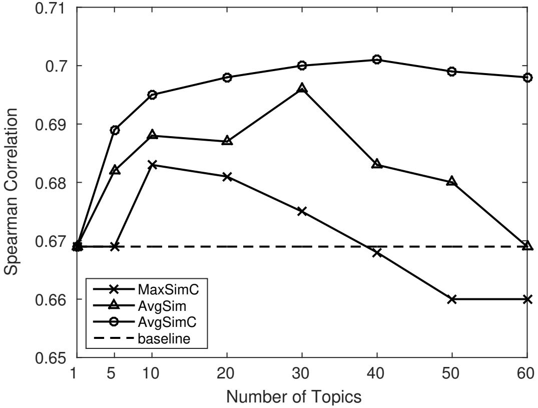
Semantic Similarity Results II

Spearman's ρ correlation for **SCWS dataset** using the **TDSMs** as a function of the number of topics, for **Web-Corpus**.



Semantic Similarity Results III

Spearman's ρ correlation for **SCWS dataset** using **TDSMs** as a function of the number of topics, for **Wikipedia**.



Affect Estimation Results

Spearman's ρ correlation for sentence affective score estimation on the SemEval 2007 Task 14 dataset.

Number of Topics	Linear Fusion	Weighted Fusion	Max Fusion
1	0.614	0.627	0.543
10	0.637	0.595	0.563
20	0.626	0.639	0.572
30	**0.646**	**0.650**	**0.603**
40	0.614	0.617	0.551
50	0.641	0.634	0.586
60	0.605	0.608	0.544

Overview

Conclusions

- Topic-based adaptation of semantic similarities
- Sub-corpora: words of interest have topic-related senses
- Linear combination of topic-specific similarities:
  state-of-the-art results on MEN dataset
  (0.838 Spearman correlation
  3.3% improvement over the state-of-the-art)
- Affect estimation with TDSMs:
  baseline (single DSM) improvement almost 4%

- Future Work
  - Optimal number of topics using semantically-driven criteria
  - Normalization, fusion of generic and topic-specific word embeddings
  - Corpora and evaluation datasets in other languages

Thank you