Mixture of Topic-based Distributional Semantic and Affective Models

Fenia Christopoulou, Eleftheria Briakou, Elias Iosif, Alexandros Potamianos
School of ECE, National Technical University of Athens, Greece

Introduction

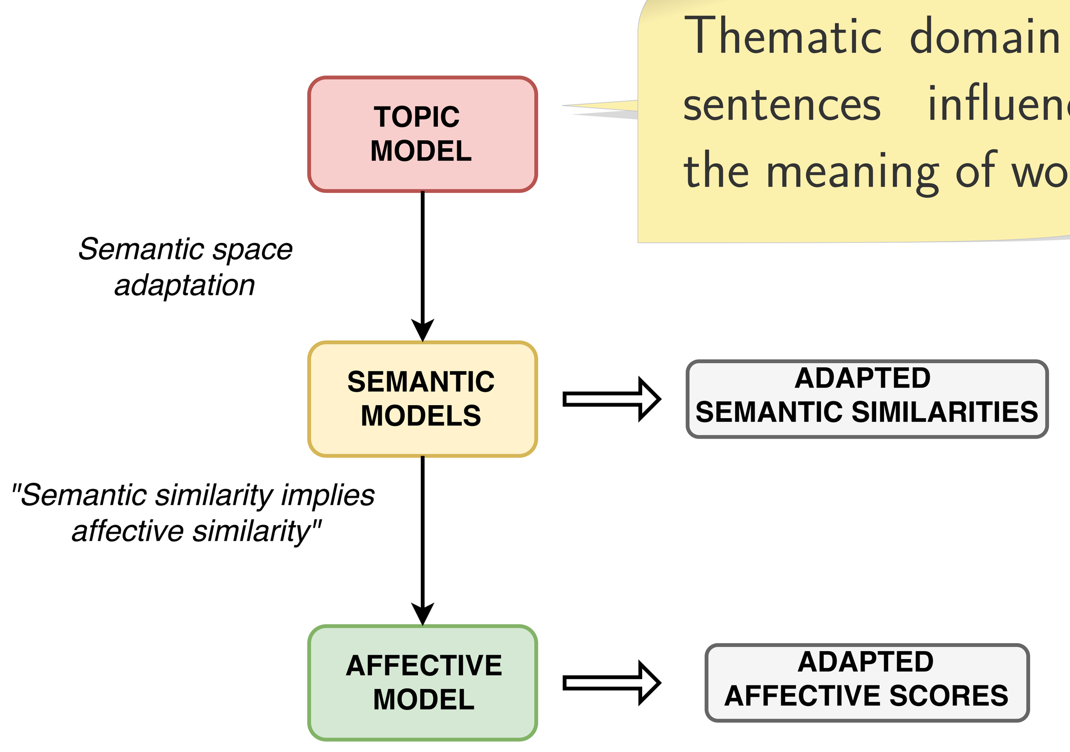
Goal: Tackle **word ambiguity** in

1. Word-level semantic similarity
2. Sentence-level affect estimation

Motivation: Traditional DSMs → 1 semantic representation with flattened senses

Overall Approach

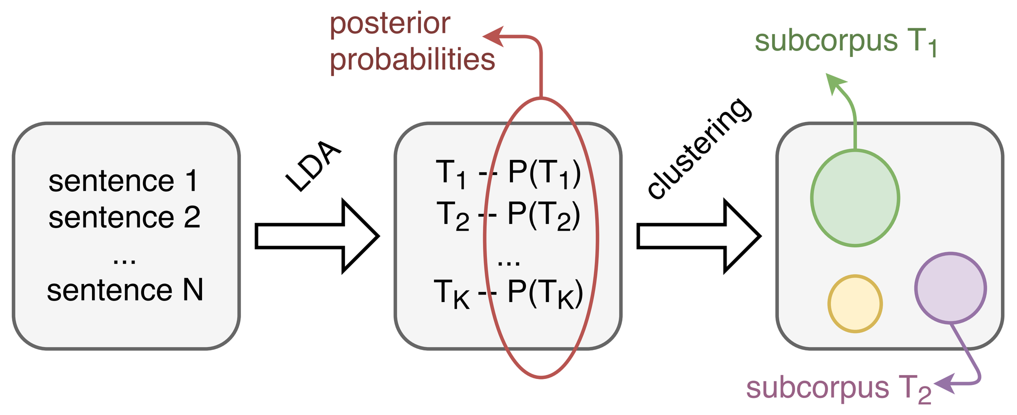
Thematic domain of sentences influences the meaning of words



Topic DSMs (TDSMs)

1. Train a probabilistic topic model
2. Apply model to the same corpus
3. Classify corpus sentences into topic-specific subcorpora

Assumption: 1 sentence contains 1 topic!



4. Train TDSMs on subcorpora
5. Estimate pairwise similarities

Similarity Computation A

- Context-independent metrics

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

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

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 DSM λ_t

- Context-dependent metrics

$c(w_i, w_j) = c(w_i) \oplus c(w_j)$: shared context of word pair

$K(c)$: candidate topics with posterior probability > 0.01

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

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

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

$p(t|c)$: posterior probability of topic t for context c

Similarity Computation B

- Fusion of topic models

Combine multiple topic-models

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

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

- Linear Regression

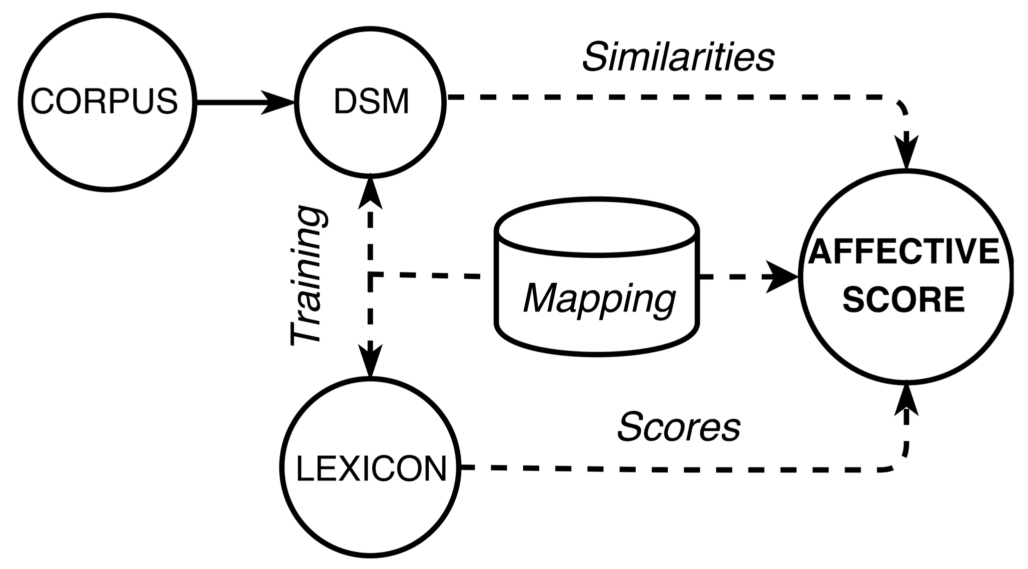
Linear combination of topic-similarities

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

β_0: bias weight; β_t: learned weight for topic t

Semantic-Affective Model

Semantic similarity implies affective similarity



- Affective space (valence)
- Words with known affective scores
- DSM trained on general-purpose corpus
- Semantic-affective space mapping [Malandrakis et al., 2011]

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

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

Topic-adapted affective scores:

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

$v_{\text{adapt}}(w_j)$: adapted affective score of word w_j; $S_{\text{AvgSimC}}(s_i, w_j; L_T)$: adapted similarity from L_T DSMs

From words to sentences (fusion):

- Linear: average of words affective scores
- Weighted: higher scores matter more
- Max: maximum absolute affective score

Data

Evaluation metric → **Spearman's ρ correlation**

- Web corpus [Iosif & Potamianos, 2015]
- Wikipedia corpus

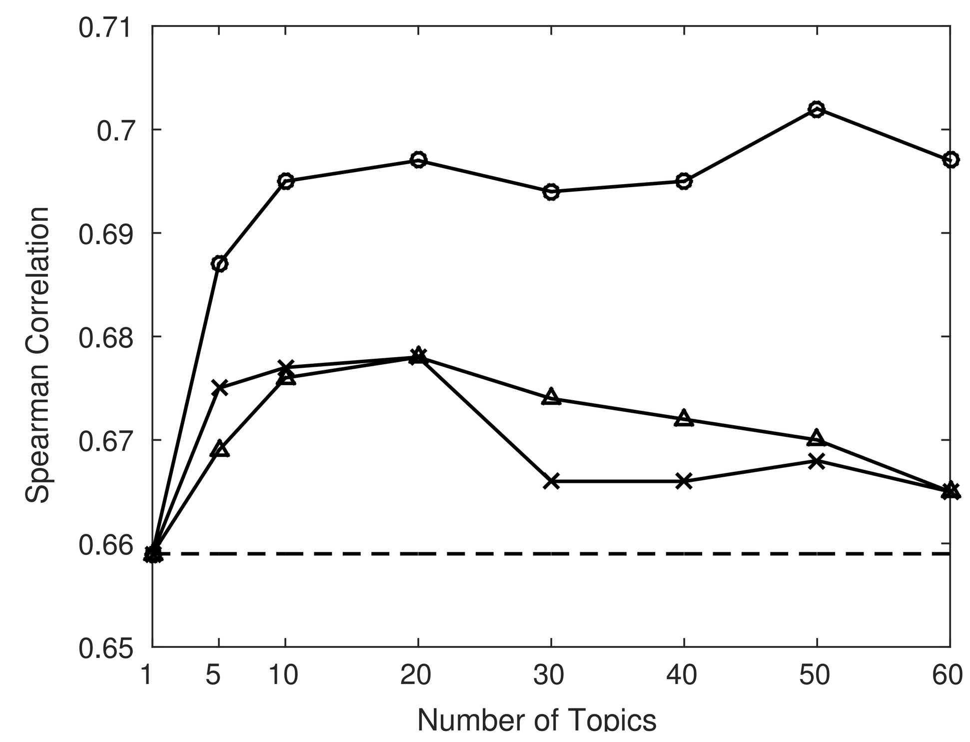
semantic similarity:	SCWS, MEN, WS-353
affective scores:	SemEval 2007 - Task 14

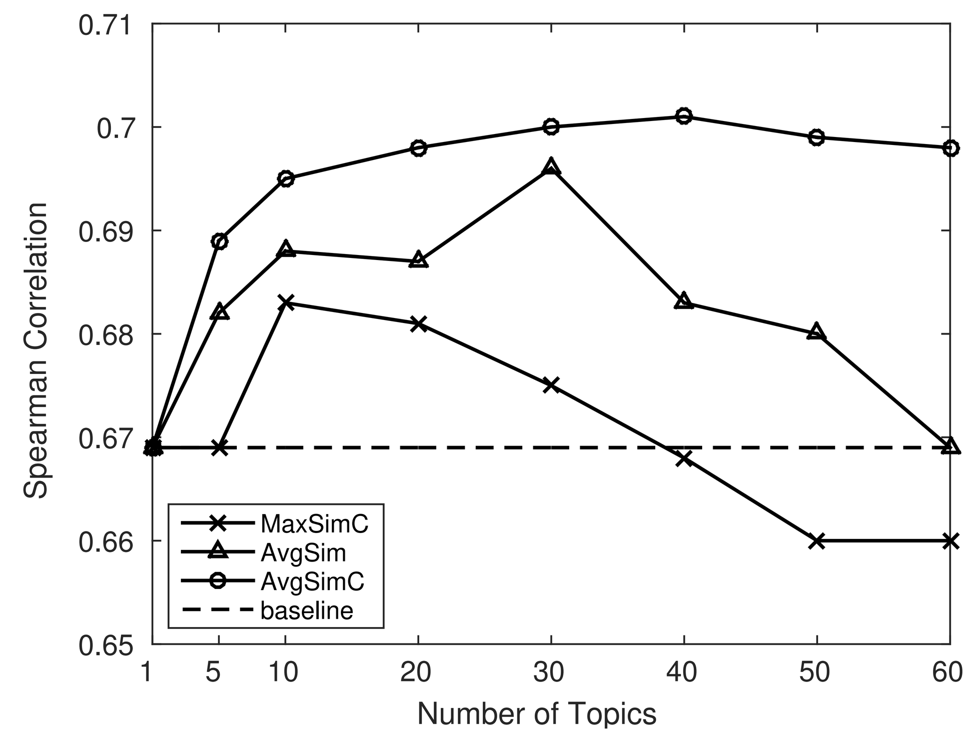
Experimental Results

- Out-of-contex datasets

Approach	WS-353	MEN
Web Corpus		
TDSMs	0.722	0.800
TDSMs-LR	0.727	0.838
No Topics	0.703	0.773
Wikipedia Corpus		
TDSMs	0.698	0.753
TDSMs-LR	0.695	0.796
No Topics	0.644	0.731

- In-context dataset (SCWS)





- Affect Estimation

|T|	Linear	Weighted	Max
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

Conclusions

- Sub-corpora where target pairs appear with topic-related senses
- Linear Regression achieves state-of-the-art results
- Affect Estimation with mixture of topic-based DSMs similarities improved by 4%

LaTeX TikZposter