Feeling is Understanding: From Affective to Semantic Spaces

Elias Iosif

School of ECE, National Technical University of Athens, Greece

"Athena" Research Center, Greece

`iosife@central.ntua.gr`

Alexandros Potamianos

School of ECE, National Technical University of Athens, Greece

"Athena" Research Center, Greece

`potam@central.ntua.gr`

Abstract

Motivated by theories of language development we investigate the contribution of affect to lexical semantics in the context of distributional semantic models (DSMs). The relationship between semantic and affective spaces is computationally modeled for the task of semantic similarity computation between words. It is shown that affective spaces contain salient information for lexical semantic tasks. We further investigate specific semantic relationships where affective information plays a prominent role. The relations between semantic similarity and opposition are studied in the framework of a binary classification problem applied for the discrimination of synonyms and antonyms. For the case of antonyms, the use of affective features results in 33% relative improvement in classification accuracy compared to the use of semantic features.

1   Introduction

Mainstream distributional semantic models (DSMs) rely solely on linguistic data, being ungrounded to the real world, i.e., features from other modalities and experiential information that are related to the acquisition of semantic knowledge are ignored. Motivated by findings from the literature of language development, according to which language acquisition is (also) grounded on communication episodes where partners exchange feelings (Tomasello et al., 2005), we consider emotion as part of lexical semantics. We argue that emotion conveys salient information, relaxing the view of emotion as "pathos" (Salovey and Mayer, 1990) that was ostracized by (traditional) models of semantics/logic.

In this paper, the affective content of words is investigated within a network-based framework regarding its contribution to lexical semantics tasks. This framework is motivated by cognitive models that rely on the distributed representation of semantic attributes (features) (Rogers and McClelland, 2004). Given a stimulus (e.g., a word), local areas (sub-spaces) are activated, triggering a number of attributes that are (semantically) related with the stimulus. The activation of attributes can be explained in the context of semantic priming according to which the presence of a word facilitates the cognitive processing of another word (McNamara, 2005). Affective priming constitutes the emotional analogue of semantic priming (Ferré and Sánchez-Casas, 2014). The key machinery of the used network is a two-tier system. The first layer constitutes a local representation scheme for encoding the semantics of target words simulating the aforementioned activation models. The activation models enable the definition of various similarity metrics in the second layer. In this work, we investigate the creation of activation models using both lexical and affective features, which are used for the computation of word semantic similarity. To the best of our knowledge this is the first computational model investigating the role of affect in semantics.

2   Related Work

Semantic similarity is the building block for numerous applications of natural language processing, such as affective text analysis (Malandrakis et al., 2013). There has been much research interest on devising data-driven approaches for estimating semantic similarity between words. Distributional semantic models (DSMs) (Baroni and Lenci, 2010) are based on the distributional hypothesis of meaning (Harris, 1954) assuming that semantic similarity between words is a function of the overlap of their linguistic contexts. DSMs can be categorized into unstructured that employ a bag-of-words model and structured that employ syntactic relationships between words (Baroni and Lenci, 2010). DSMs are typically constructed from co-occurrence statistics of word tuples that are extracted on existing corpora or on corpora specifically harvested from the web. In (Iosif and Potamianos, 2015), a language-agnostic DSM was proposed as a two-tier system motivated by cognitive considerations such as network activation and priming.. The first layer, encodes the semantics of words via the creation of lexical neighborhoods. In the second layer, similarity metrics are defined on these semantic neighborhoods. The extension of DSMs for representing the compositional aspects of lexical semantics constitutes an active research area (Baroni et al., 2014).

Analysis of text to estimate affect or sentiment is a relatively recent research topic that has attracted great interest, as reflected by a series of shared evaluation tasks, e.g., analysis of tweets (Nakov et al., 2013). Relevant applications deal with numerous domains such as news stories (Lloyd et al., 2005) and product reviews (Hu and Liu, 2004). Affective analysis is also useful for other application domains such as dialogue systems (Lee and Narayanan, 2005). Several resources enable the development of these computational models, ranging from flat lexica (e.g., General Inquirer (Stone et al., 1966) and Affective norms for English Words (Bradley and Lang, 1999)) to large lexical networks (e.g., SentiWordNet (Esuli and Sebastiani, 2006) and WordNet Affect (Strapparava and Valitutti, 2004)). Text can be analyzed for affect at different levels of granularity: from single words to entire sentences. In (Turney and Littman, 2003), the affective ratings of unknown words were predicted using the affective ratings for a small set of words (seeds) and the semantic relatedness between the unknown and the seed words. An example of sentence-level approach was proposed in (Malandrakis et al., 2013) applying techniques from n-gram language modeling.

3   Lexical Features and Metrics of Semantic Similarity

Co-occurrence-based (CC). The underlying assumption of co-occurrence-based metrics is that the co-existence of words in a specified contextual environment indicates semantic relatedness. In this work, we employ a widely-used co-occurrence-based metric, namely, Dice coefficient D (co-occurrence is considered at the sentence level).

Context-based (CT). The fundamental assumption behind context-based metrics is that *similarity of context implies similarity of meaning* (Harris, 1954). A contextual window of size $2H + 1$ words is centered on the word of interest w_i and lexical features are extracted. For every instance of w_i in the corpus the H words left and right of w_i formulate a feature vector x_i. For a given value of H the context-based semantic similarity between two words, w_i and w_j, is computed as the cosine of their feature vectors: $Q^H(w_i, w_j) = \frac{x_i.x_j}{||x_i|| \, ||x_j||}$. The elements of feature vectors can be weighted according to various schemes, while, here we use a binary scheme.

4   Affective Features and Metric of Affective Similarity

A word w is characterized regarding its affective content in a continuous (within the $[-1, 1]$ interval) space consisting of three dimensions (affective features), namely, valence (v), arousal (a), and dominance (d). For each dimension, the affective content of w is estimated as a linear combination of its' semantic similarities to a set of K seed words and the corresponding affective ratings of seeds (for the

corresponding dimension), as follows (Malandrakis et al., 2013).

$$\hat{u}(w) = \lambda_0 + \sum_{i=1}^{K} \lambda_i \, u(t_i) \, S(t_i, w), \tag{1}$$

where $t_1...t_K$ are the seed words, $u(t_i)$ is the affective rating for seed word t_i with u denoting one of the aforementioned dimensions, i.e., v, a, or d. λ_i is a trainable weight corresponding to seed t_i. $S(t_i, w)$ stands for a metric of semantic similarity (see Section 3) between t_i and w. The affective distance between two words, w_i and w_j, can be computed as the Euclidean distance over the three-dimensional space, which can be transformed into similarity.

5 Semantic and Affective Networks

In this section, we summarize the main ideas of DSMs that were proposed in (Iosif and Potamianos, 2015) for building semantic networks, which are extended here for the creation of affective networks. An overview of the semantic and affective networks is presented in Fig. 1. Each network type consists
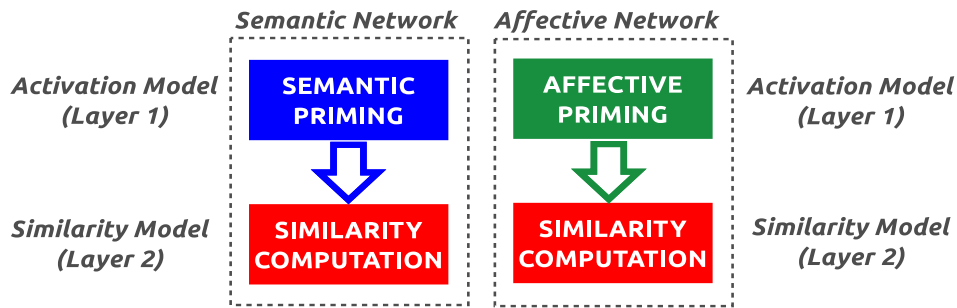

Figure 1: Overview of semantic and affective networks. Each network consists of two layers, namely, activation and similarity.

of two layers, namely, activation and similarity layer. For a target word, w_i, the first layer is used for the activation of a set of other words that are semantically/affectively related with the target. The second layer is used for the computation of semantic/affective similarity between two words for which the respective activation layers have been computed. Regardless of the network type (i.e., semantic or affective), the network is defined as an undirected (under a symmetric similarity metric) graph $F = (V, E)$ whose the set of vertices V are all words in a lexicon O, and the set of edges E contains the links between the vertices. The links (edges) between words in the network are determined and weighted according to the pairwise (semantic of affective) similarity of the vertices. For each word (target word) that is included in the lexicon, $w_i \in O$, we consider a sub-graph of F, $F_i = (N_i, E_i)$, where the set of vertices N_i includes in total n members of O, which are linked with w_i via edges E_i.

5.1 Layer 1: Activation Models

Semantic Activation Model. The computation of the semantic activation model for a target word w_i is motivated semantic priming (McNamara, 2005). The model can be represented as a F_i sub-graph, which is also referred to as the semantic neighborhood of w_i. The members of N_i (neighbors of w_i) are selected according to a semantic similarity metric (in this work, D or Q^H defined in Section 3) with respect to w_i, i.e., the n most similar words to w_i are selected. The semantic neighborhood of target w_i with size n is denoted as $L_i(n)$.

Affective Activation Model. The computation of the affective activation model for a target word w_i is motivated affective priming (Ferré and Sánchez-Casas, 2014). The model can be represented as a F_i sub-graph that denotes the affective neighborhood of w_i. The members of N_i (neighbors of w_i) are selected according to an affective similarity metric (e.g., as defined in Section 4) with respect to w_i, i.e., the n

most similar words to w_i are selected. The affective neighborhood of target w_i with size n is denoted as $A_i(n)$.

5.2 Layer 2: Similarity Model

Here, we describe two network-based similarity metrics proposed in (Iosif and Potamianos, 2015) for computing the similarity between two (target) words w_i and w_j. The metrics are defined on top of the activations models (semantic of affective) of w_i and w_j that were computed in the previous layer of the network[1].
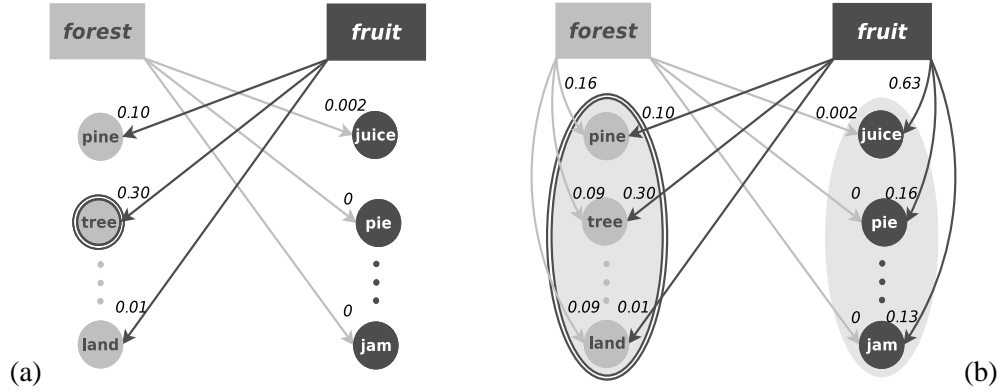

Figure 2: Example of network similarity metrics based on the activation models of two target words. The targets, "forest" and "fruit", are depicted along with their neighbors (Layer 1): {pine, tree, ..., land} and {juice, pie, ..., jam}, respectively. Arcs represent the similarities between targets and neighbors. The similarity between "forest" and "fruit" (Layer 2) is computed according to (a) maximum similarity of neighborhoods, and (b) correlation of neighborhood similarities.

Maximum Similarity of Neighborhoods. This metric is based on the hypothesis that the similarity of two words, w_i and w_j, can be estimated by *the maximum similarity of their respective sets of neighbors*, defined as follows:

$$M_n(w_i, w_j) = \max\{\alpha_{ij}, \alpha_{ji}\}, \tag{2}$$

where

$$\alpha_{ij} = \max_{x \in N_j} S(w_i, x), \quad \alpha_{ji} = \max_{y \in N_i} S(w_j, y).$$

α_{ij} (or α_{ji}) denotes the maximum similarity between w_i (or w_j) and the neighbors of w_j (or w_i) that is computed according to a similarity metric S: for semantic neighborhoods one of the metrics defined in Section 3, or the metric defined in Section 4 for affective neighborhoods. N_i and N_j are the set of neighbors for w_i and w_j, respectively. For the case of semantic neighborhoods the definition of M_n is motivated by the maximum sense similarity assumption (Resnik, 1995) hypothesizing that the most salient information in the neighbors of a word are semantic features denoting senses of this word. An example illustrating the computation of similarity between targets "forest" and "fruit" is depicted by Fig.2(a). $M_n(\text{"forest"}, \text{"fruit"}) = 0.30$ because the similarity between "fruit" and "tree" (among all neighbors of "forest") is the largest.

Attributional Neighborhood Similarity. The similarity between w_i and w_j is defined as follows:

$$R_n(w_i, w_j) = \max\{\beta_{ij}, \beta_{ji}\}, \tag{3}$$

where

$$\beta_{ij} = \rho(C_i^{N_i}, C_j^{N_i}), \quad \beta_{ji} = \rho(C_i^{N_j}, C_j^{N_j}).$$

---

[1]Similarity metrics can be applied over the semantic and affective neighborhoods of w_i and w_j. In the metric definitions we use the (generic) notations N_i and N_j to refer to the neighborhoods of w_i and w_j, respectively, regardless of the type (i.e., semantic or affective) of those neighborhoods.

$C_i^{N_i} = (S(w_i, x_1), S(w_i, x_2), \ldots, S(w_i, x_n))$ and $N_i = \{x_1, x_2, \ldots, x_n\}$. The vectors $C_j^{N_i}$, $C_i^{N_j}$, and $C_j^{N_j}$ are defined similarly as $C_i^{N_i}$. The ρ function stands for the Pearson's correlation coefficient, N_i is the set of neighbors of word w_i, and S is a similarity metric: for semantic neighborhoods one of the metrics defined in Section 3, or the metric defined in Section 4 for affective neighborhoods. The motivation behind this metric is attributional similarity, i.e., we assume that neighborhoods encode semantic or affective features of a word. Semantically/affectively similar words are expected to exhibit correlated similarities with respect to such features. The similarity computation process is exemplified in Fig.2(b) for the target words $w_i =$"forest" and $w_j =$ "fruit". The similarity vectors between the neighbors N_i of "forest" and each of the words are computed: $C_i^{N_i} = (0.16, 0.09, \ldots, 0.09)$, $C_j^{N_i} = (0.10, 0.30, \ldots, 0.01)$. Similarly, $C_i^{N_j}$, $C_j^{N_j}$ are computed for the neighbors of "fruit" and combined to estimate $R_n(\textit{"forest"}, \textit{"fruit"}) = -0.04$.

5.3 Fusion of Lexical and Affective Activation Models

In this section, we propose two schemes for the unsupervised fusion of semantic and affective activation models defined in Section 5.1. The motivation behind this idea is the hypothesis that both semantic and affective activations are triggered given lexical stimuli, e.g., the target words for which similarity is computed. In addition, for the task of similarity computation we assume that the two activation models are fused rather being exploited independently. Two types of fusion are proposed, namely, local and global. The local scheme is based on the fusion of semantic and affective neighborhoods of relatively small size. The largest possible sizes of semantic and affective neighborhoods (i.e., equal to the number of network nodes) are used for the case of global fusion.

Local. A hybrid neighborhood $N_i^{\Psi(n)}$ for a target word w_i is computed based on its lexical and affective neighborhoods, $L_i(n)$ and $A_i(n)$ of size n, as follows:

$$N_i^{\Psi(n)} = f(L_i(n), A_i(n)), \tag{4}$$

where f stands for a set operator given that $L_i(n)$ and $A_i(n)$ are represented as sets.

Global. A hybrid neighborhood $N_i^{\Omega(n)}$ of size n for a target word w_i is computed based on its lexical and affective neighborhoods, $L_i(|O|)$ and $A_i(|O|)$ of size $|O|$ (i.e., equal to the size of the lexicon O) as:

$$N_i^{\Omega(n)} = g(S(w_i, L_i(|O|)), S(w_i, A_i(|O|)); n), \tag{5}$$

where $S(w_i, L_i(|O|))$ and $S(w_i, A_i(|O|))$ stand for the vectors including the semantic and affective similarity scores between target w_i and the members of $L_i(|O|)$ and $A_i(|O|)$, respectively. Before the application of the g fusion function the two vectors should be normalized and aligned. The fusion results into a single vector of size O from which the n top-ranked values are selected and the corresponding n lexicon entries are considered as members of the neighborhood $N_i^{\Omega(n)}$.

Fusion level: function	Examples		
	Lexical model	Affective model	Fused
Local: $L_i \cup A_i$	$L_i =$\{pine, tree,...\}	$A_i =$\{taste, sugar,...\}	\{pine, tree, taste, sugar,...\}
Global: $\zeta_i^L \cdot \zeta_i^A$	$\zeta_i^L = [0.5, 0.3, \ldots]$	$\zeta_i^A = [0.2, 0.8, \ldots]$	$[0.1, 0.24, \ldots]$
Global: $\max\{\zeta_i^L, \zeta_i^A\}$	$\zeta_i^L = [0.5, 0.3, \ldots]$	$\zeta_i^A = [0.2, 0.8, \ldots]$	$[0.5, 0.8, \ldots]$

Table 1: Fusion functions for the lexical and affective activation models.

We present results for a number of simple functions for the fusion of L_i and A_i shown in Table 1. For the case of local fusion, the hybrid neighborhood is built by taking the union of semantic and affective neighborhoods. Denoting vectors $S(w_i, L_i(|O|))$ and $S(w_i, A_i(|O|))$ as ζ_i^L and ζ_i^A, respectively, two functions are used for the case of global fusion: $\zeta_i^L \cdot \zeta_i^A$ and $\max\{\zeta_i^L, \zeta_i^A\}$. The first stands for the product

of ζ_i^L and ζ_i^A. The second function gives the maximum element-wise value, i.e, for each lexicon entry and the target w_i the respective maximum semantic or affective similarity score is selected.

6   Features of Semantic Semantic Opposition

Here, we propose two feature sets that are relevant to the relations of synonymy and antonymy (also referred to as semantic opposition (Mohammad et al., 2013)). Antonymy constitutes a special lexical relation, since it embodies both the notion of (semantic) proximity and distance (Cruse, 1986). These features are based on the affective content of words and features of semantic similarity. Unlike people that can easily distinguish synonyms and antonyms, this is a challenging problem for the framework of DSMs. Both synonyms and antonyms exhibit strong associations which can be empirically verified via standard psycholinguistic experiments, as well as within the computational framework of DSMs. For example, in free association norms antonyms are frequently given as responses. Regarding DSMs, the corpus-derived statistics for synonyms and antonyms are correlated leading to comparable similarity scores. For example, in (Mohammad et al., 2013) the relatedness (similarity) scores of semantically similar (SW) and antonymous (AW) words were analyzed. Interestingly, it was found that the average score for AW was slightly higher compared to SW. The affective content of words can be considered as connotations that are added to the respective semantics. The emotional similarity between synonyms and antonyms is expected to have a contribution regarding their discrimination. For this purpose, the following features are proposed:

1) **Lex1 (lexical).** Similarity score based on direct co-occurrence counts. This can be regarded as a coefficient of semantic priming.
2) **Lex2 (lexical).** Similarity score computed according to (2) (max-based network metric). Lexical features are used for both network layers.
3) **Lex3 (lexical).** Similarity score computed according to (3) (correlation-based network metric). Lexical features are used for both network layers.
4) **Aff1 (affective).** Affective distance computed on the three-dimensional space (valence–arousal–dominance). This can be thought as a coefficient of affective priming.
5) **Aff2 (affective).** Similarity: score computed according to (2) (max-based network metric). Affective features are used for both network layers.
6) **Aff3 (affective).** Similarity score computed according to (3) (correlation-based network metric). Affective features are used for both network layers.

In essence, for each feature set (lexical and affective) there two types of similarity. The first type considers the direct similarity of the words of interest, while for the second type, the similarity is estimated via the respective neighborhoods.

7   Experiments and Evaluation Results

In this section, we investigate the role of semantic and affective features for two tasks of lexical semantics. Semantic and affective activation models are used in combination with the aforementioned network-based similarity metrics for the computation of word semantic similarity. This is presented in Section 7.1, while the fusion of the two activation types is shown in 7.2. In Section 7.3, semantic and affective features are evaluated in the framework of semantic opposition. This is done as a binary classification problem for the discrimination of synonyms and antonyms.

7.1   Word Semantic Similarity Computation

Creation of Networks. A lexicon consisting of $8,752$ (single-word) English nouns was taken from the SemCor3 corpus. For the extraction of the textual features a web harvested corpus was created as follows.

For each lexicon entry an individual query was formulated and the $1,000$ top ranked results (document snippets) were retrieved using the Yahoo! search engine and aggregated. The affective ratings (v, a and d) for these nouns were computed using as seeds the manually annotated ANEW lexicon (Bradley and Lang, 1999) (600 seeds were used) and estimating the λ weights of (1) according to (Malandrakis et al., 2013). Regarding $S(.)$ used in (1), the context-based (CT) similarity metric exploiting text features was applied. The network creation consisted of two main steps: 1) computation of semantic and affective neighborhoods as described in Section 5, 2) computation of similarity scores using M_n and R_n defined by (2) and (3), respectively. For the case of semantic neighborhoods two types of similarity metrics (in conjunction with the respective textual features) were applied: co-occurrence-based (CC), and context-based (CT) with $H = 1$.

Evaluation. The task of noun semantic similarity computation was used for evaluation purposes with respect to the following datasets (i) MC (Miller and Charles, 1998),(ii) RG (Rubenstein and Goodenough, 1965), and (iii) WS353 (Finkelstein et al., 2002), retaining those pairs that were included in the network. The Pearson's correlation coefficient against human ratings was used as evaluation metric.

| Type of feature for		Network-based metric	Number of neighbors (n)				
Selection of neighbors (1st layer)	Similarity computation (2nd layer)		10	30	50	100	150
MC dataset							
Lexical (CC)	Lexical (CT)	M_n	0.48	0.80	0.83	**0.91**	0.90
Lexical (CT)	Lexical (CC)	R_n	0.83	0.78	0.80	0.78	0.76
Affective	*Lexical (CC)*	R_n	0.85	**0.91**	0.88	0.85	0.83
RG dataset							
Lexical (CC)	Lexical (CT)	M_n	0.57	0.74	0.78	**0.86**	0.82
Lexical (CT)	Lexical (CC)	R_n	0.65	0.71	0.72	0.72	0.72
Affective	*Lexical (CC)*	R_n	0.78	**0.80**	0.79	0.77	0.74
WS353 dataset							
Lexical (CC)	Lexical(CT)	M_n	0.42	0.55	0.59	0.64	**0.65**
Lexical (CT)	Lexical(CC)	R_n	0.63	0.58	0.59	0.56	0.55
Affective	*Lexical (CC)*	R_n	0.63	**0.68**	**0.68**	0.65	0.63

Table 2: Correlation for word similarity computation.

The performance for various neighborhood sizes is presented in Table 2 for two approaches regarding the activation model (Layer 1) followed by the neighborhood-based similarity estimation (Layer 2). Two types of activation models are used for the computation neighborhoods, namely, lexical and affective. Once the neighborhoods are computed, the network metrics M_n and R_n are employed for the similarity computation based on lexical features. Overall, there are two basic settings: *Lexical+Lexical* and *Affective+Lexical*. The core novelty of this work is on the exploitation of affective features for the activation model, i.e., the Affective+Lexical approach. For the sake of completeness, the results when using textual features only (Lexical+Lexical) are presented for the respective best performing metrics and feature types (according to (Iosif and Potamianos, 2015)): CC/CT for M_n and CT/CC for R_n. Regarding the Affective+Lexical approach, the performance is reported only for R_n that was found to outperform the (omitted) M_n metric. It is notable[2] that the Affective+Lexical combination performs very well being competitive[3] against the best Lexical+Lexical approach, as well as other state-of-the-art approaches (Agirre et al., 2009). Specifically, the Affective+Lexical combination achieves higher (0.68 vs. 0.65)

---

[2]This was experimentally verified using the affective word ratings given by human annotators (ANEW affective lexicon (Bradley and Lang, 1999)), instead of the automatically estimated ratings produced by (1).

[3]The detailed comparison of the proposed affective models with other lexical DSMs is beyond the scope of this study.

and equal (0.91) correlation scores -compared to the Lexical+Lexical combination- for the WS353 and MC datasets, respectively. The Affective+Lexical combination consistently achieves higher (or equal) performance compared to both Lexical+Lexical combinations when few (10-50) neighbors are used.
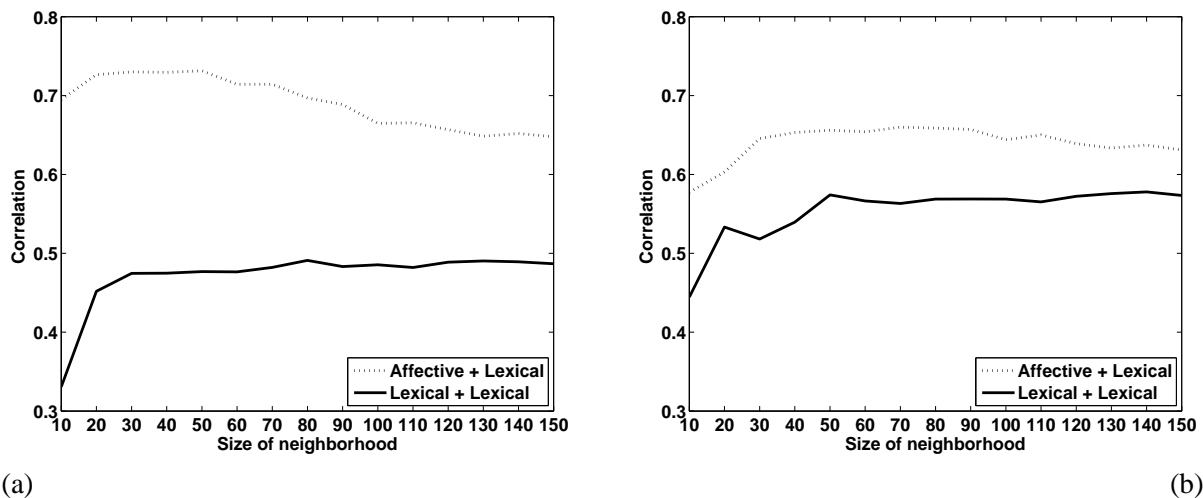
(a)   (b)

Figure 3: Correlation for word similarity computation as a function of neighborhood size for pairs consisting of words with: (a) distant affective magnitude (150 pairs from WS353), and (b) comparable affective magnitude (122 pairs from WS353). Results are shown for Lexical+Lexical (solid line) and Affective+Lexical (dotted line) approaches.

Motivated by the very good performance of the Affective+Lexical approach, we conducted further investigation regarding the role of affective information with respect to the affective relation of the words for which the similarity is computed. For this purpose, the pairs of the largest experimental dataset (WS353) where distinguished into two groups according to the affective magnitude of their constituents words. The first group includes pairs whose both constituents have high or low affective magnitude (i.e., words with comparable magnitude), e.g., (king, queen). The remaining pairs were included in the second group (i.e., words with distant magnitude), e.g., (psychology, depression). The discrimination resulted into 122 and 150 pairs consisting of words with comparable and distant affective magnitude, respectively. The performance of the Lexical+Lexical and Affective+Lexical approaches using the R_n similarity metric is shown as a function of the neighborhood size in Fig. 3(a) for words with distant affective magnitude, and in Fig. 3(b) for words with comparable affective magnitude. We observe that the Affective+Lexical approach consistently achieves higher correlation compared to the Lexical+Lexical approach for both groups. The superiority of the Affective+Lexical approach is shown more clearly for the case of words with distant affective magnitude (Fig. 3(a)).

7.2   Fusion of Lexical and Affective Activation Models

Fusion level	Fusion function	Number of neighbors				
		10	30	50	100	150
Best individual model		0.63	**0.68**	**0.68**	0.65	0.63
Best lexical model		0.42	0.55	0.59	0.64	0.65
Local	$L_i \cup A_i$	0.45	0.47	0.44	0.47	0.46
Global	$\zeta_i^L \cdot \zeta_i^A$	0.46	0.48	0.50	0.49	0.48
	$\max\{\zeta_i^L, \zeta_i^A\}$	0.63	**0.68**	**0.68**	0.65	0.63

Table 3: Correlation for word similarity computation (WS353 dataset).

In this section, the evaluation results for the fusion of semantic and affective models (Layer 1) are presented. The fusion schemes shown in Table 1 were used for the computation of hybrid neighbor-

169

Semantic relation	Baseline (random)	Feature types	
		Lexical (Lex1,Lex2,Lex3)	Affective (Aff1,Aff2,Aff3)
Synonymy	50%	61%	62%
Antonymy	50%	61%	**82%**

Table 4: Classification accuracy for synonymy and antonymy: lexical vs. affective feature sets.

Semantic relation	Baseline (random)	Lexical features			Affective features		
		Lex1	Lex2	Lex3	Aff1	Aff2	Aff3
Synonymy	50%	51%	61%	59%	61%	61%	51%
Antonymy	50%	55%	61%	61%	**81%**	**82%**	50%

Table 5: Classification accuracy for synonymy and antonymy for individual lexical and affective features.

hoods. The network-based similarity metric R_n was applied over the hybrid neighborhoods for the computation of semantic similarity between words (Layer 2). The performance is presented in Table 3 for the largest dataset (WS353) with respect to various neighborhood sizes. The correlation achieved by the best performing individual model (Affective+Lexical using R_n) is included for comparison purposes. The performance of the best model based solely on lexical features (Lexical+Lexical using M_n) is also presented. Regarding the different fusion schemes, the highest performance is obtained for the global approach using the maximum-based function ($\max\{\zeta_i^L, \zeta_i^A\}$). This scheme yields performance that is identical to the best individual model. Also, we observe that the best fusion scheme consistently outperforms the Lexical+Lexical approach for $10 - 100$ neighbors.

7.3 Synonymy vs. Antonymy

Here, we compare the performance of semantic and affective features (described in Section 6) for the discrimination of word pairs the fall into two categories, synonyms and antonyms. The word pairs were taken from two sets of WordNet synonyms and opposites[4] We retained those pairs that were included in the networks described in Section 7.1. In total, 172 pairs are contained in each category for a total of 344 pairs. The experimental dataset include pairs such as (happiness, felicity) and (comedy, tragedy) that correspond to synonyms and antonyms, respectively. Support Vector Machines[5] with linear kernel were applied for classification. For evaluation purposes, 10-fold cross validation (10-FCV) was used, while classification accuracy was used as evaluation measurement.

The classification accuracy is shown for each category in Table 4 with respect to two feature sets: 1) all lexical features (Lex1–Lex3), and 2) all affective features (Aff1–Aff3)[6]. The baseline performance (yielded by random classification) is also presented. Both features types exceed the baseline for synonyms and antonyms. The main observation is that the set of affective features outperforms the lexical feature set for the case of antonyms, i.e., 82% vs. 61% classification accuracy. Regarding synonyms, lexical and affective features yield almost identical performance. The moderate discrimination ability of lexical features was expected since both synonyms and antonyms exhibit high similarity scores as measured in the framework of DSMs. These observations suggest that the affective information is a major contributor for the case of antonyms, which is not surprising since such words are emotionally distant. The performance for all individual features in presented in Table 5 for each category. It is observed that the similarities based on word co-occurrence (Lex1) give the lowest performance for both synonyms and antonyms, while the network-based similarities (Lex2 and Lex3) yield slightly higher results. The key observation is that the top performance, i.e., greater than 80%, can be achieved either using the simple

---

[4]http://www.saifmohammad.com/WebPages/ResearchInterests.html\#Antonymy.

[5]Similar results were obtained with other classifiers, e.g., Naive Bayes.

[6]For the network metrics we used $n = 30$, however, similar results were achieved for other values of n, e.g., 10, 50, 100.

affective similarity (Aff1) or the maximum-based network similarity metric (Aff2). Given the lack of a standard dataset for this task, the comparison of different DSMs is not easy. A corpus-based algorithm was evaluated with respect to similar a task (synonym/antonym discrimination for 136 pairs[7]) achieving 75% classification accuracy under 10-FCV (Turney, 2011).

8   Conclusions

The affective spaces were shown to contain salient information for estimating semantic similarity. The Affective+Lexical approach achieved competitive performance compared to (an example of) the mainstream paradigm of distributional semantic models (i.e., the Lexical+Lexical approach). Moreover, the affective models were found to be more appropriate for the first network layer (compared to the lexical models) when the words for which similarity is computed exhibit distant affective magnitude. To the best of our knowledge, this is the first empirical indication that the affect can be regarded as another source of information that plays a role for the task of semantic similarity estimation between words. Correlation-based similarity metrics and smaller neighborhoods were shown to perform better for Affective+Lexical DSMs. Another major finding is that the affective features are superior to the lexical ones for the case of antonym identification. Regarding the fusion of lexical and affective activation models, the global scheme (i.e., across the entire network) was found to outperform the local one. Further research is needed for understanding the complementarities of affective and semantic spaces, which is important for the design of improved fusion schemes. Last but not least, the role of affective features should be investigated with respect to more semantic tasks (e.g., paraphrasing) and other types of semantic relations and linguistic phenomena (e.g., figurative language).

References

Agirre, E., E. Alfonseca, K. Hall, J. Kravalova, M. Pasca, and A. Soroa (2009). A study on similarity and relatedness using distributional and WordNet-based approaches. In *Proc. of NAACL: HLT*, pp. 19–27.

Baroni, M., R. Bernardi, and R. Zamparelli (2014). Frege in space: A program for compositional distributional semantics. *Linguistic Issues in Language Technologies 9*(6), 5–110.

Baroni, M. and A. Lenci (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics 36*(4), 673–721.

Bradley, M. and P. Lang (1999). Affective norms for English words (ANEW): Stimuli, instruction manual and affective ratings. Tech. report C-1. The Center for Research in Psychophysiology, Univ. of Florida.

Cruse, D. A. (1986). *Lexical Semantics*. Cambridge University Press.

Esuli, A. and F. Sebastiani (2006). SentiWordNet: A publicly available lexical resource for opinion mining. In *Proc. of Conference on Language Resources and Evaluation*, pp. 417–422.

Ferré, P. and R. Sánchez-Casas (2014). Affective priming in a lexical decision task: Is there an effect of words' concreteness? *Psicológica 35*, 117–138.

Finkelstein, L., E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppin (2002). Placing search in context: The concept revisited. *ACM Transactions on Information Systems 20*(1), 116–131.

---

[7]Including verbs, while our network includes only nouns.

Harris, Z. (1954). Distributional structure. *Word 10*(23), 146–162.

Hu, M. and B. Liu (2004). Mining and summarizing customer reviews. In *Proc. of Conference on Knowledge Discovery and Data Mining*, KDD '04, pp. 168–177.

Iosif, E. and A. Potamianos (2015). Similarity computation using semantic networks created from web-harvested data. *Natural Language Engineering 21*(01), 49–79.

Lee, C. M. and S. S. Narayanan (2005). Toward detecting emotions in spoken dialogs. *IEEE Transactions on Speech and Audio Processing 13*(2), 293–303.

Lloyd, L., D. Kechagias, and S. Skiena (2005). Lydia: A system for large-scale news analysis. In *Proc. SPIRE*, Number 3772 in Lecture Notes in Computer Science, pp. 161–166.

Malandrakis, N., A. Potamianos, E. Iosif, and S. Narayanan (2013). Distributional semantic models for affective text analysis. *IEEE Transactions on Audio, Speech, and Language Processing 21*(11), 2379–2392.

McNamara, T. P. (2005). *Semantic priming: Perspectives from Memory and Word Recognition*. Psychology Press.

Miller, G. and W. Charles (1998). Contextual correlates of semantic similarity. *Language and Cognitive Processes 6*(1), 1–28.

Mohammad, S. M., B. J. Dorr, G. Hirst, and P. D. Turney (2013). Computing lexical contrast. *Computational Linguistics 39*(3), 555–590.

Nakov, P., S. Rosenthal, Z. Kozareva, V. Stoyanov, A. Ritter, and T. Wilson (2013). Semeval 2013 task 2: Sentiment analysis in twitter. In *Proc. of Second Joint Conference on Lexical and Computational Semantics (*SEM), Seventh International Workshop on Semantic Evaluation*, pp. 312–320.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxanomy. In *Proc. of International Joint Conference for Artificial Intelligence*, pp. 448–453.

Rogers, T. T. and J. L. McClelland (2004). *Semantic Cognition. A Parallel Distributed Processing Approach*. The MIT Press.

Rubenstein, H. and J. B. Goodenough (1965). Contextual correlates of synonymy. *Communications of the ACM 8*(10), 627–633.

Salovey, P. and J. D. Mayer (1990). Emotional intelligence. *Imagination, Cognition and Personality 9*(3), 185–211.

Stone, P. J., D. C. Dunphy, M. S. Smith, and D. M. Ogilvie (1966). *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Strapparava, C. and A. Valitutti (2004). WordNet Affect: An affective extension of WordNet. In *Proc. of Conference on Language Resources and Evaluation*, pp. 1083–1086.

Tomasello, M., M. Carpenter, J. Call, T. Behne, and H. Moll (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences 28*, 675–691.

Turney, P. D. (2011). Analogy perception applied to seven tests of word comprehension. *Journal of Experimental and Theoretical Artificial Intelligence 23*(3), 343–362.

Turney, P. D. and M. L. Littman (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems 21*(4), 315–346.