Associative and Semantic Features Extracted From Web-Harvested Corpora

Elias Iosif[*], Maria Giannoudaki[*], Eric Fosler-Lussier[†], Alexandros Potamianos[*]

[*] Dept. of Electronics & Computer Engineering, Technical University of Crete, Chania 73100, Greece
{iosife,maria,potam}@telecom.tuc.gr
[†] Dept. of Computer Science & Engineering, Ohio State University, Columbus, OH 43210, USA
fosler@cse.ohio-state.edu

Abstract

We address the problem of automatic classification of associative and semantic relations between words, and particularly those that hold between nouns. Lexical relations such as synonymy, hypernymy/hyponymy, constitute the fundamental types of semantic relations. Associative relations are harder to define, since they include a long list of diverse relations, e.g., "Cause-Effect", "Instrument-Agency". Motivated by findings from the literature of psycholinguistics and corpus linguistics, we propose features that take advantage of general linguistic properties. For evaluation we merged three datasets assembled and validated by cognitive scientists. A proposed priming coefficient that measures the degree of asymmetry in the order of appearance of the words in text achieves the best classification results, followed by context-based similarity metrics. The web-based features achieve classification accuracy that exceeds 85%.

Keywords: associative relations, semantic relations, priming

1. Introduction

We address the problem of automatic classification of associative and semantic relations between words, and particularly those that hold between nouns. Lexical relations such as synonymy, hypernymy/hyponymy, constitute the fundamental types of semantic relations (Cruse, 1986). Associative relations are harder to define, since they include a long list of diverse relations, e.g., "Cause-Effect", "Instrument-Agency". From the perspective of cognitive scientists, associative relatedness is triggered by the co-occurrence of words (McNamara, 2005), while the definition of semantic relatedness is controversial. The boundary between semantic and associative relations is not always clear, since highly associated words tend to be semantically related, e.g., (cat,dog).

Previous research efforts have investigated semantic relations, such as the identification of synonyms, (Iosif and Potamianos, 2010), hyponyms, (Caraballo, 1999). Also, the identification of other relations has attracted the research interest, e.g., the Task 8 of SemEval'10 dealt with the classification of various relations (Hendrickx et al., 2010). To our knowledge there have been very few computational efforts for the discrimination between associative and semantic relations, e.g., (Turney, 2008).

Such classification can be beneficial for a wide range of language technologies. For example, in statistical language modeling, class-based language models (Brown et al., 1992) have long been used to extend the coverage of the model – words in classes should typically be semantically related (i.e., sister hyponyms of the same hypernym). However, trigger models (Lau et al., 1993) try to find words that change the probability distribution over other words, which is more of an associative relationship (e.g., postman – letter). Other technologies might use relationships in a different way: spoken dialogue systems often have an ontology of semantically related concepts (which one can attempt to learn from corpus data (Pargellis et al., 2004)); query expansion techniques for information retrieval have also utilized semantically related concepts (Fang, 2008). On the other hand, information extraction tasks may benefit from knowing associative relationships between words, since the contextual information leading to a decision to extract some piece of information is more likely to be associative in nature.

We propose an automated computational approach that discriminates between associative and semantic relations. Text-based lexical and hit-based features are extracted from the web in order to classify given pairs of concepts as semantic or associative. These features do not rely on manually selected syntactic patterns, such as Hearst's patterns for the identification of "is-a" relations and semantic role labeling, but are rather motivated by general cognitive and linguistic principles. Specifically, we propose two novel features: (a) the degree of priming (co-occurrence asymmetry) as a function of the distance between the two words in text, and (b) the rate of change of context-based lexical similarity as a function of the context window size. Evaluation proceeds on a dataset containing 238 associative and semantic relations, which they were appropriately assembled by cognitive scientists in order to exclude any fuzzy relations.

2. Related Work

Semantic similarity metrics can be divided into two broad categories: (i) metrics that rely on knowledge resources, and (ii) corpus- or web-based metrics that do not require any external knowledge source. A representative example of the first category are metrics that exploit the WordNet ontology (Miller, 1990). For computing the similarity between words these metrics incorporate features such as the length of paths between the two words (Resnik, 1995; Jiang and Conrath, 1997) or the information content of their least subsumer, estimated from a corpus (Wu and Palmer, 1994; Leacock and Chodorow, 1998). WordNet glosses are also used as features in (Patwardhan and Pedersen., 2006). A study that reviews in depth the major WordNet-based metrics is provided in (Budanitsky and Hirst, 2006). Corpus-based metrics usually extract contextual features from text for computing semantic similarity. Web-based methods employ search engines to estimate the frequency of word co-occurrence (Vitanyi, 2005; Gracia et al., 2006; Turney, 2001) or construct corpora (Bollegala et al., 2007; Iosif and Potamianos, 2010). The identification and extraction of other types of relations has been mainly studied through the use of linguistic patterns. Lexico-syntactic patterns were applied in the influential work of Hearst (Hearst, 1992), for the identification of hyponymy, followed by numerous similar approaches, e.g., (Caraballo, 1999). Pattern-based approaches were also employed for the meronymy relation (Girju et al., 2003).

3. A Review of Web-based Metrics

In this section, we review two types of web-based metrics (Iosif and Potamianos, 2010): web-page counts (hits) for computing relatedness between words, and contextual features for computing semantic similarity.

3.1. Hit-based metrics

The underlying assumption of hit-based metrics is that two words that co-exist in the same document are related. Sets of documents $\{D\}$ returned by a search engine with query words $w_i, \ldots, w_{i+n}$ are notated by $\{D; w_i, \ldots, w_{i+n}\}$; the cardinality is noted by $\mid D; w_i, \ldots, w_{i+n} \mid$. We investigate four different metrics in this work:

Jaccard coefficient (J): computes the relatedness between w_i and w_j by employing the overlap of document sets in which they appear, i.e.,

$$J(w_i, w_j) = \frac{\mid D; w_i, w_j \mid}{\mid D; w_i \mid + \mid D; w_j \mid - \mid D; w_i, w_j \mid} \qquad (1)$$

The Jaccard coefficient takes values between 0 (total dissimilarity) and 1 (absolute similarity).

Dice coefficient (C): a variation of the Jaccard coefficient, also takes values in $[0, 1]$. It is defined as:

$$C(w_i, w_j) = \frac{2 \mid D; w_i, w_j \mid}{\mid D; w_i \mid + \mid D; w_j \mid}. \qquad (2)$$

Mutual information (I): (as defined in (Bollegala et al., 2007)) computes the similarity between w_i and w_j by calculating the mutual dependence of the random variables W_i and W_j that represent the number of documents indexed by w_i and w_j, respectively. It is defined as the point-wise mutual information between W_i and W_j:

$$I(w_i, w_j) = \log \frac{\frac{|D; w_i, w_j|}{|D|}}{\frac{|D; w_i|}{|D|} \frac{|D; w_j|}{|D|}} \qquad (3)$$

According to this info-theoretic measure, if two variables are independent, the knowledge of the one variable does not provide any information about the other, and their mutual information equals to 0.

Google-based Semantic Relatedness (G): The Normalized Google Distance is a dissimilarity metric proposed in (Vitanyi, 2005; Cilibrasi and Vitanyi, 2007). It is defined as:

$$E(w_i, w_j) = \frac{\max\{L\} - \log \mid D; w_i, w_j \mid}{\log \mid D \mid - \min\{L\}}, \qquad (4)$$

where $L = \{\log \mid D; w_i \mid, \ \log \mid D; w_j \mid\}$. This metric is unbounded, assigning dissimilarity scores that range from 0 to ∞. The Normalized Google Distance was adopted in (Gracia et al., 2006) in order to propose a bounded (in $[0, \ 1]$) metric, called Google-based Semantic Relatedness, defined as:

$$G(w_i, w_j) = e^{-2E(w_i, w_j)}. \qquad (5)$$

3.2. Context-based metrics

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 v_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:

$$S^H(w_i, w_j) = \frac{v_i . v_j}{\mid\mid v_i \mid\mid \ \mid\mid v_j \mid\mid}. \qquad (6)$$

The elements of feature vectors are weighted according to two schemes: 1) Binary (B), and 2) Log of Term Frequency (LTF). B assigns 1 to a feature if it appears within the context of w_i, otherwise 0. According to the LTF scheme, the weight assigned to a contextual feature is a function of the logarithm of its frequency, normalized by the logarithm of the frequency of w_i. For more details consult (Iosif and Potamianos, 2010).

4. Associative and semantic features

In this section, we propose two novel features for discriminating between associative and semantic relations using information automatically extracted from the web. Syntactic patterns are also investigated as features.

4.1. Hit-based priming coefficient

Hit-based metrics (summarized in Section 3.1.) employ co-occurrence counts without taking into account: (i) the order of appearance of each word, and (ii) the distance (i.e., the number of words that intervene) between occurrences of the two words. Next, we motivate the use of these features for classifying associative and semantic relations.

The use of word order is motivated by findings in cognitive science and psycholinguistics, about the asymmetry of the priming phenomenon with respect to word pairs. In psycholinguistics, the notion of priming refers to the cognitive processing that takes place when two words in a certain order are presented to a human subject. In this framework, the first word p ("prime") serves as a stimulus that facilitates (or primes) the cognitive processing of the second word t ("target") (McNamara, 2005). The selection of prime and target is determined experimentally for each word pair based on human response time, where response time is assumed to be inversely proportional to the strength of priming (or relatedness). Once the prime and target are defined, their usual order (p, t) is known as "forward", while the reverse order (t, p) is called "backward". It has been found that the difference between forward and backward priming is statistically significant for many related word pairs, e.g., responses to the pair ('light','bulb') were reported to be quicker than the responses to the pair ('bulb','light') (McNamara, 2005; Koriat, 1981). Similar observations regarding the asymmetry of order of appearance within co-occurrence were also reported in the NLP literature (Church and Hanks, 1990). However, data related to this phenomenon have been analyzed without further exploration of the cognitive aspects of the problem.

Our goal is to define a "priming coefficient", i.e., a single metric that characterizes the degree of asymmetry in the forward and backward co-occurrence counts. Since priming is sensitive to ordering, we compute"forward" and "backward" co-occurrence counts (as a function of the distance between words) for each word pair. We expect that word pairs (p,t) with strong priming should appear much more often in the forward rather than the backward order. We expect priming to be a good discriminator between associative and semantic relations as psycholinguistics have suggested that priming effects can be of different magnitude for these different relations (Plaut, 1995; Ferrand and New, 2003).

Instead of using raw co-occurrence counts to estimate the priming coefficient, we propose to use the normalized hit-based metrics defined in Section 3.1. We introduce a variation of hit-based metrics that computes separately forward and backward co-occurrence counts, conditioned on the distance d between words. For a word pair (w_i, w_j), the forward relatedness $R_{f,m}$ is defined as

$$R^A_{f,m}(w_i, w_j) = A(w_i, w_j; d = m), \tag{7}$$

computed only for forward co-occurrence counts with distance d that is equal to m words. Function $A(.)$ denotes any of the hit-based metric defined in Section 3.1. Similarly, backward relatedness is defined as:

$$R^A_{b,m}(w_i, w_j) = A(w_j, w_i; d = m). \tag{8}$$

Total relatedness Λ^A_m is defined as the sum of the forward and backward relatedness

$$\Lambda^A_m(w_i, w_j) = R^A_{f,m}(w_i, w_j) + R^A_{b,m}(w_i, w_j) \tag{9}$$

for metric A, word pair (w_i, w_j) and distance equal to m. Finally, the priming coefficient Ψ^A_m is defined as the normalized absolute difference between forward and backward relatedness

$$\Psi^A_m(w_i, w_j) = \frac{|\, R^A_{f,m}(w_i, w_j) - R^A_{b,m}(w_i, w_j)\,|}{R^A_{f,m}(w_i, w_j) + R^A_{b,m}(w_i, w_j)}. \tag{10}$$

The priming coefficient is equal to 0 when the forward and backward co-occurrence counts are equal (no priming) and 1 when a word pair only appears with the forward (or backward) order (very strong priming).

4.2. Slope of text-based similarity

In Section 3.2., a context-based metric was defined that has been used in the literature for estimating the strength of semantic relations between words. In general, the strength of both semantic and associative relations covers a wide range from weak to strong; as a result, the relation strength by itself is a poor discriminator of the semantic vs associative class.

Based on observations in psycholinguistics (Ferrand and New, 2003) and computational linguistics (Hearst, 1992), words that are semantically similar, especially synonyms and words that belong to the same semantic class, can be identified by lexico-syntactic patterns from their immediate vicinity. For this case, context-based semantic similarity metrics are also shown to better correlate with human judgements when small contextual windows are used to compute similarity (Iosif and Potamianos, 2010). Associative relations often imply a shared pragmatic context that is also evident from lexical similarity in the not-so-immediate vicinity. Thus, the relevance of lexical features extracted from context is expected to be a function of the contextual window size.

According to the above considerations, we assume that the migration from syntactic to pragmatic features by increasing the size of H, will affect differently the context similarity of associative and semantic relations. For this purpose, we compute the difference of semantic similarity scores across different sizes of H. In particular, we focus on window sizes that differ exactly by one (first-order differences). Consider two words w_i and w_j. The difference of their similarity scores with respect to window sizes, H_x and H_y, is computed as:

$$S^{H_x}_{H_y}(w_i, w_j) = S^{H_x}(w_i, w_j) - S^{H_y}(w_i, w_j), \quad (11)$$

for $H_x - H_y = 1$. The similarities $S^{H_x}(w_i, w_j)$ and $S^{H_y}(w_i, w_j)$ are computed according to (6).

4.3. Linguistic patterns

We also examine whether specific syntactic patterns can discriminate between associative and semantic relations. By manual inspection of our data we have summarized the most common patterns for associative ([A1],[A2]) and semantic ([S1],[S2]) relations, respectively:

[A1] Complex Noun Phrases (NPs): $[NP_{term1|term2}[NP_{term1|term2}]]$, e.g., "**Ocean wave** energy is captured directly from surface waves or from pressure fluctuations."

[A2] Terms co-occurring in argument positions: $[NP_{term1}[VP[NP_{term2}]]]$, e.g., "...why do **giraffes** have long **necks**..."

[S1] The two terms in coordinative constructions: $[NP_{term1}]$ AND|OR $[NP_{term2}]$, e.g., "**Beet and radish** roots are similar in shape, but beets are usually larger than radishes."

[S2] The two terms in extended coordinative constructions, involving one additional NP between the NPs of interest: $[NP_{term1|term2}]$, $[NP]$ AND|OR $[NP_{term1|term2}]$, e.g., "... professional **carpet, upholstery and rug** cleaners in the Chicago ... "

Overall, associative noun pairs are expected to surface as arguments of the same phrase: in pattern A1 one NP is contained into the other, while in pattern A2 both NPs are manifested in the argument positions of the same VP (subject and object of the verb). Semantically related noun pairs form NPs that are structurally independent of each other; when they co-occur in close proximity they are usually connected with conjunctions.

5. Experimental Dataset

There are relatively few datasets containing rated associative or semantic relations between word pairs or terms, most of them containing fewer than 50 pairs. Lack of a standardized dataset of adequate size is a barrier to computational approaches that require fair amounts of data for training and testing. In this work, we have merged datasets taken from three different studies from the literature of psycholinguistics (Chiarello et al., 1990; Perea and Gotor, 1997; Ferrand and New, 2003) for a total of 238 relations, equally split between 119 associative and 119 semantic relations (Table 1). All three datasets were designed for psycholinguistic experiments related to priming, and contain only "pure" associative and semantic relations, avoiding word pairs that lie in the boundaries of the two relations. Well-established lists of free association norms, e.g., (Palermo and Jenkins, 1964; Nelson et al., 1998), were used for the selection of "pure" associatively related pairs. Such lists are constructed by collecting the responses of human subjects when stimuli words are presented to them and they are asked to give the very first word they recall. Regarding "pure" semantic relations, the relevant pairs were selected according to the following criteria: (i) the words of each pair are members of the same semantic category and they have high similarity score, and (ii) they are not included in lists of free association norms. The similarity score incorporated by the first criterion typically is estimated by collecting similarity ratings given by subjects.

Dataset No	# of semantic rel.	# of associative rel.
1	42	42
2	48	48
3	29	29
Total	119	119

Table 1: Experimental datasets.

The semantically related pairs in datasets 1 and 3 consist exclusively of words that belong to the same semantic category, i.e., hyponyms of the same hypernym. The semantically related pairs in dataset 2 consist of words with various degrees of synonymy. Some indicative examples of the relations included in the experimental datasets are presented in Table 2.

Dataset No	Semantic rel.	Associative rel.
1	brass–iron	onion–tears
1	velvet–linen	hammer–nail
1	bacon–steak	pilot–plane
2	boat–ship	board–wood
2	work–labor	nucleus–center
2	fume–steam	hour–clock
3	clarinet–flute	drill–hole
3	pancake–waffle	cow–milk
3	rug–carpet	suitcase–trip

Table 2: Examples of dataset relations.

6. Experimental procedure

We compute hit-based metrics and text-based metrics through web search engines as described below:

6.1. Hit-based metrics

The number of word co-occurrences is estimated by Yahoo! search API[1] that returns the number of web hits given a particular query. We wish to compute the number of hits for the word pair (w_i, w_j) under the following constraints (i) w_i precedes w_j, and (ii) their distance, defined in Section 4.1. as the number of intervening words, is equal to m, i.e., $m = 2$. This is achieved by the query "$w_i \star \star w_j$" for $m = 2$. The "$\star$" symbol is a special search metacharacter, matching any word (Bollegala et al., 2010). Using this query formulation, we retrieve the number of hits for both forward and backward ordering of the words up to a particular distance m. Once the number of hits is retrieved, the total relatedness $\Lambda_m^A(w_i, w_j)$ is computed according to (9), for each of the hit-based metrics A. Similarly, the priming coefficient $\Psi_m^A(w_i, w_j)$ is computed according to (10). For each word pair, $\Lambda_m^A(w_i, w_j)$ and $\Psi_m^A(w_i, w_j)$ are computed, using the four hit-based metrics $A = \{J, C, I, G\}$ and for distance values $m = 0, \ldots, 10$.

6.2. Text-based metrics

For the computation of text-based semantic similarity between the words of associative and semantic relations, we need to build a corpus from the web. For each word pair (w_i, w_j), we download 1000 snippets of web documents using the Yahoo! Search API. The web search is performed according to the conjunctive query "w_i AND w_j", ensuring that both words co-occur in the same snippet, for reasons explained in (Iosif and Potamianos, 2010). Once the snippets are retrieved, we compute for each word pair: (i) the semantic similarity score, $S^H(w_i, w_j)$, according to (6), and (ii) the difference of similarities across different window sizes, $S_{H_y}^{H_x}(w_i, w_j)$, according to (11). The similarities are computed using B and LTF weighting schemes (Section 3.2.) for contextual window sizes $H = 1, \ldots, 10$.

7. Evaluation Results

In this section, we present results for associative vs semantic relation classification using the dataset described in Section 5. We used the support vector machine classifier provided by Weka[2]; similar results were obtained using naive Bayes classifier (not reported here due to lack of space). The evaluation was performed according to a 10-fold validation procedure. The evaluation results are reported in terms of classification accuracy.

In Fig. 1(a), the classification accuracy is shown for: (i) total relatedness, $\Lambda_m^J(w_i, w_j)$, computed according to (9) (solid line), and (ii) priming coefficient, $\Psi_m^J(w_i, w_j)$, computed according to (10) (dotted line). Classification accuracy is plotted as a function of m, the distance between words. It is clear that total relatedness achieves very poor accuracy that lies close to chance. The poor performance at $m = 0$ is an indication that the asymmetry of priming at the bigram level can not discriminate associative and semantic relations. The priming coefficient obtains good accuracy around 80% for most values of m, excluding the value $m = 1$. The discriminative ability of the priming coefficient improves for distance values around 6 or 7 (although the differences in performance are not statistically significant). In Table 3, the classification precision is summarized for a number of hit-based metrics for the total relatedness and priming coefficient. These results were obtained by joining the individual features for distances $m = 0, \ldots, 10$ into a single vector. Again, significantly higher results, up to 86.5%, are achieved by the priming coefficient. There is no significant difference among the hit-based metrics.

Hit-based metrics	Accuracy	
	Total related.	Priming coef.
J	53.2%	86.5%
C	52.7%	86.5%
I	56.5%	85.7%
G	62.9%	86.5%

Table 3: Classification accuracy for total relatedness and priming coefficient.

In Fig. 1(b), the classification accuracy as a function of the window size H is shown for: (i) context-based similarity $S^H(w_i, w_j)$ computed according to (6) (solid line), and (ii) similarity slope $S_{H_y}^{H_x}(w_i, w_j)$ computed according to (11) (dotted line). For both of them, the binary B weighting scheme was used. Context-based similarity $S^H(w_i, w_j)$ is shown to be a relatively poor discriminator of associative vs semantic relations, and the achieved accuracy remains low, $55\% - 62\%$, for all values of H. The similarity slope $S_{H_y}^{H_x}(w_i, w_j)$ metric also performs poorly with the exception of window $H = 2$; performance for $S_{H=1}^{H=2}(w_i, w_j)$ exceeds 70% accuracy. Classification accuracies for both metrics $S^H(w_i, w_j)$, $S_{H_y}^{H_x}(w_i, w_j)$ and for both B and LTF weighting schemes are presented in Table 4. Results are computed for the joined feature vector containing values computed for contextual window sizes $H = 1, \ldots, 10$. For comparison, we have also included the accuracy for three WordNet-based similarity metrics, namely Leacock-Chodorow (Leacock and Chodorow, 1998), Resnik (Resnik, 1995), and Vector (Patwardhan and Pedersen., 2006). These metrics were computed us-

[1] `http://developer.yahoo.com/search/`
[2] `http://www.cs.waikato.ac.nz/ml/weka/`

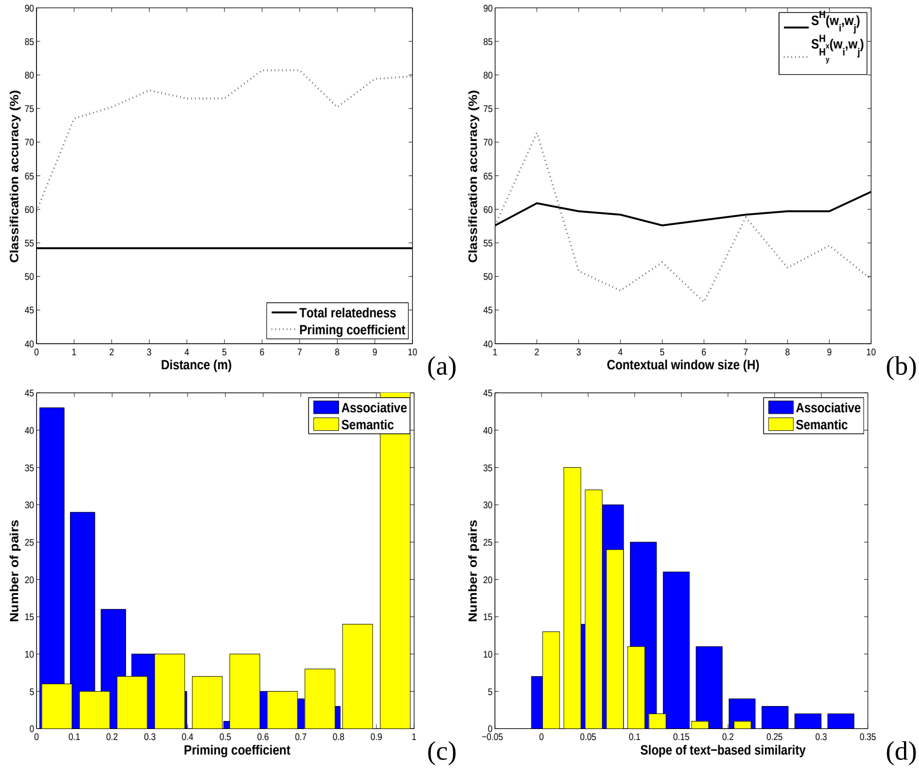

Figure 1: Classification accuracy for: (a) total relatedness $\Lambda_m^J(w_i, w_j)$, and priming coefficient $\Psi_m^J(w_i, w_j)$ as a function of distance m for the Jaccard (J) hit-based metric, (b) semantic similarity $S^H(w_i, w_j)$ and slope $S_{H_y}^{H_x}(w_i, w_j)$ metrics as a function of the window size H, using the binary B weighting scheme. Histograms for associative and semantic pairs: (c) priming coefficient $\Psi_5^J(w_i, w_j)$, (d) similarity slope $S_{H=1}^{H=2}(w_i, w_j)$.

Metrics of semantic similarity	Accuracy
$S^H(w_i, w_j)$, B scheme	62.6%
$S^H(w_i, w_j)$, LTF scheme	62.6%
$S_{H_y}^{H_x}(w_i, w_j)$, B scheme	**71.8%**
$S_{H_y}^{H_x}(w_i, w_j)$, LTF scheme	64.3%
WN: Leacock-Chodorow	71.0%
WN: Resnik	**75.6%**
WN: Vector	54.2%

Table 4: Accuracy for context-based similarity, similarity slope and WordNet-based (WN) similarity metrics.

ing the WordNet::Similarity package, developed by Pedersen and it is freely available through CPAN [3]. The $S^H(w_i, w_j)$ similarity metrics achieve relatively low accuracy, below 63%. WordNet-based metrics display diverse performance ranging from 54.2% for the Vector metric to 75.6% for the Resnik metric. The accuracy achieved by the slope $S_{H_y}^{H_x}(w_i, w_j)$ metric is up to 71.8% for the B weighting scheme.

To further investigate the behaviour of the best performing features, we have plotted their histograms for associative and semantic word pairs. In Fig. 1(c), we show the histogram for the priming coefficient $\Psi_5^J(w_i, w_j)$. The priming coefficient for the associative relations tends to be lower than that of semantic relations, especially for larger values of distance m. The histograms of the values of $S_{H=1}^{H=2}(w_i, w_j)$ metric are shown in Fig. 1(d). Both histograms have positive means, i.e., context-based semantic similarity increases when going from window size one to size two. However, the increase for associative relations is higher.

We have also combined the best performing features: (i) Ψ_m^G priming coefficient using the G hit-based metric, and (ii) $S_{H_y}^{H_x}(w_i, w_j)$ text-based metric using B scheme, by simply taking the union of their feature sets. This combination achieved slightly higher accuracy of 87.8%. Finally, we report results separately on dataset 2 that contains synonyms as semantic pairs, and compare the results with datasets 1 and 3. The results are presented in Table 5. Note that the accuracy drops for dataset 2 (synonyms) for both the priming coefficient and, especially, the similarity slope. This is an indication that synonyms might be harder to separate from associative pairs; however, due to the limited size of dataset 2 (29 assoc. and 29 sem. relations) no general conclusions can be drawn.

[3] `http://search.cpan.org/`

Features	Set 1, 3	Set 2	All sets
$\Psi_m^G(w_i, w_j)$	87.7%	82.8%	86.5%
$S_{H_y}^{H_x}(w_i, w_j)$	76.1%	56.9%	71.8%
Both features			87.8%

Table 5: Accuracy for datasets 1, 3 vs dataset 2.

Also, some preliminary results on the classification between semantic and associative relations using linguistic patterns (on the same web corpus) are provided. The most accurate pattern for associative relations is A1 (complex NPs) achieving classification accuracy of 66%. For semantic relations the S1 pattern with coordinative constructions performs better, although its performance is below 60%. When all four patterns were used classification accuracy of up to 73.5% is achieved.

Last, in order to further validate our best performing feature, Ψ_m^G, we used some types of relations taken from the field of semantic role labeling, assuming that they can serve as associative ones. Regarding semantic relations we retained the relations of dataset 1. In particular, we considered four distinct types of relations taken from the SemEval2010–Task 8, "Multi-Way Classification of Semantic Relations Between Pairs of Nominals" (Hendrickx et al., 2010): (i) "Cause–Effect", (ii) "Instrument–Agency", (iii) "Component–Whole", and (iv) "Member–Collection". For each type of the above relations, we created a distinct dataset including the semantic relations of dataset 1 and an equal number of randomly selected examples. For each dataset we evaluated the proposed priming coefficient regarding the classification of semantic and associative relations. For all datasets the classification accuracy is very similar and exceeds 80%, even for medium values of distance ($m = 3$). These results provide an additional confirmation regarding the good performance of the proposed feature, while they are consistent with the results obtained for the datasets assembled by cognitive scientists.

8. Conclusions

Motivated by findings in the psycholinguistics and computational linguistics literature, we investigated the problem of automatically classifying relations between words into either associative or semantic, using information extracted from the web. Two new features were proposed designed specifically for this classification task, namely, the priming coefficient measuring the asymmetry in the order of appearance of the word pair and the first-order difference (slope) of the context-based semantic similarity with respect to the contextual window size. For associative relations the priming coefficient takes significantly smaller values as the distance between the two words increases, while for semantic relations priming is less affected by word distance. For words that are semantically related their contextual similarity is higher for immediate rather than for distant context (small vs. large contextual windows); for associative relations context similarity is less affected by window size. The priming coefficient is shown to be a good feature for discriminating between the two classes, achieving classification accuracy up to 86%. The slope of the contextual similarity achieves good classification results, up to 72% accuracy. Overall, we have shown that it is feasible to classify associative and semantic relations without using lexical or syntactic patterns, but rather general linguistic properties measured through lexical corpus statistics, e.g., order of appearance, co-occurrence, distance, contextual similarity. We make available [4] a resource containing more than 9.000 priming coefficients, computed for the pairs of the experimental datasets.

Further research is needed with larger datasets to verify the universality of these claims. Also special cases of associative and semantic relations should be investigated and the relative performance of the proposed features should be evaluated. The proposed features could be also relevant for investigating the differences between various types of semantic relationships, as well as for studying the priming phenomenon across different languages within the proposed computational framework.

9. Acknowledgements

Elias Iosif and Maria Giannoudaki were partially funded by the Basic Research Programme, Technical University of Crete, Project Number 99637: "Unsupervised Semantic Relationship Acquisition by Humans and Machines: Application to Automatic Ontology Creation".

10. References

D. Bollegala, Y. Matsuo, and M. Ishizuka. 2007. Measuring semantic similarity between words using web search engines. In *Proc. of International Conference on World Wide Web*, pages 757–766.

D. Bollegala, Y. Matsuo, and M. Ishizuka. 2010. Relational duality: Unsupervised extraction of semantic relations between entities on the web. In *Proc. International World Wide Web Conference*, pages 151–160.

P. Brown, P. deSouza, R. Mercer, V. Della Pietra, and J. Lai. 1992. Class-based n-gram models of natural language. *Computational Linguistics*, 18(4):467–479.

A. Budanitsky and G. Hirst. 2006. Evaluating WordNet-based measures of semantic distance. *Computational Linguistics*, 32:13–47.

S. A. Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proc. of the annual meeting of the Association for Computational Linguistics: HLT*, pages 120–126.

[4] `http://www.telecom.tuc.gr/~iosife/downloads.html`

C. Chiarello, C. Burgess, L. Richards, and A. Pollock. 1990. Semantic and associative priming in the cerebral hemispheres: Some words do, some words don't ... sometimes, some places. *Brain and Language*, 38(1):75–104.

K. W. Church and P. Hanks. 1990. Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1):22–29.

R. L. Cilibrasi and P. Vitanyi. 2007. The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, 19(3):370–383.

D. A. Cruse. 1986. *Lexical Semantics*. Cambridge University Press.

H. Fang. 2008. A re-examination of query expansion using lexical resources. In *Proc. of the annual meeting of the Association for Computational Linguistics*, pages 139–147.

L. Ferrand and B. New. 2003. Semantic and associative priming in the mental lexicon. In P. Bonin, editor, *Mental lexicon: Some words to talk about words*, pages 25–43. Hauppauge, NY: Nova.

R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proc. Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, pages 1–8.

J. Gracia, R. Trillo, M. Espinoza, and E. Mena. 2006. Querying the web: A multiontology disambiguation method. In *Proc. of International Conference on Web Engineering*, pages 241–248.

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

M. A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proc. of Conference on Computational Linguistics*, pages 539–545.

I. Hendrickx, S. N. Kim, Z. Kozareva, P. Nakov, D. Ó Séaghdha, S. Padó, M. Pennacchiotti, L. Romano, and S. Szpakowicz. 2010. Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In *Proc. of the 5th International Workshop on Semantic Evaluation*, pages 33–38.

E. Iosif and A. Potamianos. 2010. Unsupervised semantic similarity computation between terms using web documents. *IEEE Transactions on Knowledge and Data Engineering*, 22(11):1637–1647.

J. Jiang and D. Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. of International Conference on Research on Computational Linguistics*, pages 19–33.

A. Koriat. 1981. Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition*, 9:587–598.

R. Lau, R. Rosenfeld, and S. Roukos. 1993. Trigger-based language models: a maximum entropy approach. In *Proc. International Conference on Acoustics, Speech and Signal Processing*, pages 45–48.

C. Leacock and M. Chodorow. 1998. Combining local context and WordNet similarity for word sense identification in WordNet. In C. Fellbaum, editor, *An Electronic Lexical Database*, pages 265–283. MIT Press.

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

G. Miller. 1990. Wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–312.

D. L. Nelson, C. L. McEvoy, and T. A. Schreiber. 1998. The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.

D. S. Palermo and J. J. Jenkins. 1964. *Word Association Norms: Grade School Through College*. University of Minessota Press.

A. Pargellis, E. Fosler-Lussier, C. H. Lee, A. Potamianos, and A. Tsai. 2004. Auto-induced semantic classes. *Speech Communication*, 43(3):183–203.

S. Patwardhan and T. Pedersen. 2006. Using WordNet-based context vectors to estimate the semantic relatedness of concepts. In *Proc. of the EACL Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together*, pages 1–8.

M. Perea and A. Gotor. 1997. Associative and semantic priming effects occur at very short stimulus-onset asynchronies in lexical decision and naming. *Cognition*, 62(2):223–240.

D. C. Plaut. 1995. Semantic and associative priming in distributed attractor network. In *Proc. of 17th Annual Conference of the Cognitive Science Society*, pages 37–42.

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

P. D. Turney. 2001. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proc. of the European Conference on Machine Learning*, pages 491–502.

P. D. Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proc. of International Conference on Computational Linguistics*, pages 905–912.

P. Vitanyi. 2005. Universal similarity. In *Proc. of Information Theory Workshop on Coding and Complexity*, pages 238–243.

Z. Wu and M. Palmer. 1994. Verbs semantics and lexical selection. In *Proc. of the annual meeting on Association for Computational Linguistics*, pages 133–138.