BabelRelate! A Joint Multilingual Approach to Computing Semantic Relatedness

Roberto Navigli and **Simone Paolo Ponzetto**
Dipartimento di Informatica
Sapienza Università di Roma
Rome, Italy
–navigli,ponzetto˝@di.uniroma1.it

Abstract

We present a knowledge-rich approach to computing semantic relatedness which exploits the joint contribution of different languages. Our approach is based on the lexicon and semantic knowledge of a wide-coverage multilingual knowledge base, which is used to compute semantic graphs in a variety of languages. Complementary information from these graphs is then combined to produce a 'core' graph where disambiguated translations are connected by means of strong semantic relations. We evaluate our approach on standard monolingual and bilingual datasets, and show that: i) we outperform a graph-based approach which does not use multilinguality in a joint way; ii) we achieve uniformly competitive results for both resource-rich and resource-poor languages.

Introduction

Over recent years research in semantic technologies has had a major impact by enabling and improving a wide range of web-based applications, such as search (Egozi, Markovitch, and Gabrilovich 2011), clustering web search results (Navigli and Crisafulli 2010), collaborative content management (Krötzsch et al. 2007), as well as category discovery of queries (Reisinger and Paşca 2011) and videos (Toderici et al. 2010), to name just a few. Complementary to this trend, the multilingual nature of the Web has been a major driving force behind research on multilingual text processing – cf. the development of cross-lingual semantic retrieval models like, for example, those proposed by Dumais, Landauer, and Littman (1996), Potthast, Stein, and Anderka (2008) and Cimiano et al. (2009). However, even if these methods demonstrate the beneficial effects of semantics for cross-lingual web applications, they are each distributional in nature. As such, despite being based on sound statistical approaches and thus able to exploit large repositories of text like the Web, they typically have to rely on flat representations such as vectors and their algebraic properties. Moreover, they usually work on different languages separately, i.e. they project each language of interest onto a different semantic space and perform operations on these vector spaces.

In this paper we explore a knowledge-based approach to computing semantic relatedness across different languages, a task originally introduced by Hassan and Mihalcea (2009).

Computing semantic relatedness using lexical knowledge resources is a well-explored line of research (Budanitsky and Hirst 2006): however, its application to a multilingual scenario has so far been hampered by the lack of wide-coverage multilingual knowledge bases. In this study we leverage knowledge from the Web and compute semantic relatedness using BabelNet (Navigli and Ponzetto 2010), a very large multilingual lexical knowledge resource which integrates an online encyclopedia (Wikipedia[1]) with a computational lexicon (WordNet, Fellbaum (1998)). Thanks to the fact that it is a multilingual semantic network, BabelNet provides us with: i) structured representations, i.e. graphs, for concepts and their relations; ii) lexical knowledge about word senses and their translations into different languages. We use these two characteristics synergistically in order to exploit the information from different languages in a unified way and improve the performance of computing semantic relatedness on every one of them. Given a pair of words in two languages we use BabelNet to collect their translations, compute semantic graphs in a variety of languages, and then combine the empirical evidence from these different languages by intersecting their respective graphs. Our results indicate that *jointly exploiting multiple languages*: i) *improves the performance for all languages* over an approach which does not use multilingual information in a joint way; ii) enables us to *achieve state-of-the-art performance on both resource-rich and resource-poor languages*.

Joint multilingual semantic relatedness

Our method takes as input a pair of words in two languages and a multilingual lexical knowledge base[2] (i.e. a graph whose nodes are concepts lexicalized in different languages and whose edges express semantic relations between concepts). It then computes a semantic relatedness score which indicates how strongly related the input words are on the basis of the lexico-semantic information they convey. We compute these scores by means of a novel algorithm, whose pseudocode is presented in Algorithm 1 and which consists of four main steps:

[1] http://www.wikipedia.org

[2] We use BabelNet as our reference knowledge base, however our algorithm can be used with any multilingual lexical knowledge resource providing adequate lexicographic coverage.

Algorithm 1 Joint multilingual semantic relatedness.

Input: a set of two word-language pairs $\{(w_1, l_1), (w_2, l_2)\}$
a set of languages L
BabelNet BN
Output: a semantic relatedness score $sr \in [0, 1]$

1: $G_{base} \leftarrow getGraph((w_1, l_1), (w_2, l_2), BN)$

2: $translations_1 \leftarrow \emptyset$
3: $translations_2 \leftarrow \emptyset$
4: **for each** $l_i \in L - \{l_1\}$
5: $\quad translation_1 \leftarrow translate(w_1, l_1, l_i)$
6: $\quad translations_1 \leftarrow translations_1 \cup \{(translation_1, l_i)\}$
7: **for each** $l_i \in L - \{l_2\}$
8: $\quad translation_2 \leftarrow translate(w_2, l_2, l_i)$
9: $\quad translations_2 \leftarrow translations_2 \cup \{(translation_2, l_i)\}$

10: $graphs \leftarrow \emptyset$
11: **for each** $(translation_1, l) \in translations_1$
12: $\quad$ **for each** $(translation_2, l') \in translations_2$
13: $\quad\quad graphs \leftarrow graphs \cup getGraph((translation_1, l), (translation_2, l'), BN)$

14: $G_{joint} \leftarrow G_{base}$ $\qquad \triangleright\ G_{joint} = (V_{joint}, E_{joint})$
15: **if** $V_{joint} = \emptyset$ **then**
16: $\quad G_{joint} \leftarrow \underset{G \in graphs}{\mathrm{argmax}}\ score_{CO}(G)$
17: $\quad graphs \leftarrow graphs - \{G_{joint}\}$
18: **for each** $G \in graphs$ $\qquad \triangleright\ G = (V, E)$
19: $\quad V' \leftarrow V_{joint} \cap V$
20: $\quad E' \leftarrow E_{joint} \cap E$
21: $\quad$ **if** $V' \neq \emptyset$ **then**
22: $\quad\quad V_{joint} \leftarrow V'$
23: $\quad\quad E_{joint} \leftarrow E'$

24: $sr \leftarrow 0$
25: **for each** $s_1 \in Senses_{\mathrm{BN}}(w_1, l_1)$
26: $\quad$ **for each** $s_2 \in Senses_{\mathrm{BN}}(w_2, l_2)$
27: $\quad\quad G'_{joint} \leftarrow subgraph(G_{joint}, s_1, s_2)$
28: $\quad\quad score \leftarrow score_{SR}(G'_{joint}, s_1, s_2)$
29: $\quad\quad$ **if** $score > sr$ **then**
30: $\quad\quad\quad sr \leftarrow score$

31: **return** sr

1. Semantic graph construction (line 1). We start by selecting the subgraph of BabelNet which contains the paths between senses of the input words w_1 and w_2. We do this by building a labeled directed graph $G = (V, E)$ following the procedure of Navigli and Lapata (2010), which connects possible senses of w_1 with senses of w_2:

i) We first define the set V of nodes of G as the set of all Babel synsets containing w_1 and w_2, i.e., $V := Senses_{\mathrm{BN}}(w_1) \cup Senses_{\mathrm{BN}}(w_2)$. Initially, the set of edges of G is empty, i.e., $E := \emptyset$.

ii) Next, we connect the nodes in V using the paths found between them in BabelNet. Formally, for each vertex $v \in V$, we perform a depth-first search along the BabelNet graph and every time we find a node $v' \in V$ ($v \neq v'$) along a simple path $v, v_1, \ldots, v_k, v'$, we add all intermediate nodes and edges of this path to G, i.e., $V := V \cup \{v_1, \ldots, v_k\}$, $E := E \cup \{(v, v_1), \ldots, (v_k, v')\}$.

The result is a subgraph of BabelNet consisting of (1) the Babel synsets corresponding to the senses of the input words, (2) all intermediate synsets along all paths that connect them. We call this the *semantic graph* of w_1 and w_2.

2. Creating multilingual semantic graphs (lines 2–13). We next create semantic graphs for all those languages L other than the input ones. First, we obtain translations for each input word by collecting from BabelNet the lexicalizations of its senses in each language $l \in L$ (lines 2–9). In order to focus the construction of the multilingual semantic graphs on those senses which are predominant across different languages, the function *translate* returns for each language the most frequent translation of the input word in BabelNet: this consists of the synonym in the language of interest that occurs most frequently within all Babel synsets containing the input word. For instance, the word bank occurs in, among others, the following Babel synsets: (a) { bank_{EN}, banca_{IT}, …, banco_{ES} }; (b) { bank_{EN}, $\text{salvadanaio}_{\text{IT}}$, …, hucha_{ES} }; (c) { bank_{EN}, banca_{IT}, …, banco_{ES} }. Thus, its Italian and Spanish most frequent translations are banca and banco, respectively. Next, for each pair of translations, we construct a semantic graph following the same procedure used for the input words (lines 10–13). The result is a set of graphs, each containing the paths between senses of the input words' translations found in BabelNet.

3. Semantic graph intersection (lines 14–23). We then combine the information from different languages to find the 'core' subgraph of BabelNet which connects senses of the input words, as well as of their translations. Our hunch here is that intersecting the graphs of the input words and their translations will help us filter out implausible paths due to spurious translations and weak semantic relations. To build the core graph G_{joint}, we start with the semantic graph obtained from the input words w_1 and w_2 (line 14). If this graph is empty, i.e., no connecting path could be found between the input words in BabelNet, we select the highest scoring graph from the set *graphs* of those obtained from the input words' translations (based on the scoring function $score_{CO}$, see below), set G_{joint} to it (line 16), and remove this graph from *graphs* (line 17). To find the 'best' graph we need to define the scoring function $score_{CO}$. All nodes and relations being equal *a priori*, we want the graph which is semantically more consistent (i.e., which does not contain loose semantic relations) and enables the graphs from other languages to provide their contribution (i.e., it has the highest probability of yielding a non-empty intersection as output). Accordingly, we score our semantic graphs using a function which combines the graph size with a global measure of graph compactness $CO(G)$ (Botafogo, Rivlin, and Shneiderman 1992), namely its degree of cross-referencing:

$$score_{CO}(G) = |V| \times CO(G) .$$

The compactness of a graph $G = (V, E)$ is computed as:

$$CO(G) = \frac{Max - \sum_{u \in V} \sum_{v \in V} d(u, v)}{Max - Min} ,$$

where $d(u, v)$ is the length of the shortest path between nodes u and v, $Max = K|V|(|V| - 1)$ is the maximum value of the distance sum (for a completely disconnected graph) and $Min = |V|(|V|-1)$ is its minimum value (for a fully connected graph). K is a constant value expressing the distance between two nodes if there is no path connecting them: here, we follow Navigli and Lapata (2010) and set K to $|V|$, since the length of any shortest path is less than $|V|$. Once we have this initial graph G_{joint} we iterate over each graph $G \in graphs$, i.e. the set of graphs generated from the translations into the other languages (line 18). At each step we intersect the set of nodes and edges of G_{joint} with those of G (lines 19–20) and update the former if the node intersection is not empty (lines 21-23).

4. Semantic relatedness computation (lines 24–30). In the final phase we use the intersection graph from the previous step to compute semantic relatedness. To this end we assign to a word pair the relatedness score of those senses which maximize the relatedness scoring function, i.e.:

$$score_{SR}(w_1,w_2) = \max_{\substack{s_1 \in Senses_{\mathrm{BN}}(w_1),\\ s_2 \in Senses_{\mathrm{BN}}(w_2)}} score_{SR}(G,s_1,s_2)\ ,$$

where s_1 and s_2 are Babel synsets containing any of the senses of the respective input words w_1 and w_2, and G is a semantic graph containing paths found between them in BabelNet. We define the relatedness scoring function $score_{SR}$ as follows. Previous work on computing semantic relatedness using lexical resources has developed a wide variety of different measures which are all, even if to different extents, developed for hierarchical structures (i.e., *is-a* relations within a taxonomy such as WordNet). In contrast, most of the relations found in BabelNet are topically-associative, non-hierarchical ones from Wikipedia. Thus, in our scenario we cannot apply measures based on notions such as the depth of the taxonomy (e.g., Leacock and Chodorow (1998)) or the least common subsumer (Wu and Palmer (1994) and Resnik (1999), *inter alia*). For this reason, as well as that of being able to quantify the contribution of jointly exploiting multiple languages in a simple setting (i.e., leaving out any additional performance gain obtained from more complex measures), we adopt a simple node counting scheme:

$$score_{SR}(G,s_1,s_2) = \max_{p \in paths(G,s_1,s_2)} \frac{1}{length(p)}\ ,$$

where $paths(G,s_1,s_2)$ is the set of all possible paths between s_1 and s_2 in graph G, and $length(p)$ is the number of nodes in path p. We apply this measure as follows. We initialize the relatedness score sr to 0 (line 24) and iterate over each pair of senses s_1 and s_2 of the input words (lines 25–30). At each iteration we first select the subgraph G'_{joint} of G_{joint} which contains all paths between s_1 and s_2 (line 27), use it to compute the relatedness score (line 28), and update the highest relatedness score so far (lines 29-30). Finally, as a result of the execution of the algorithm, the relatedness score sr is returned (line 31).

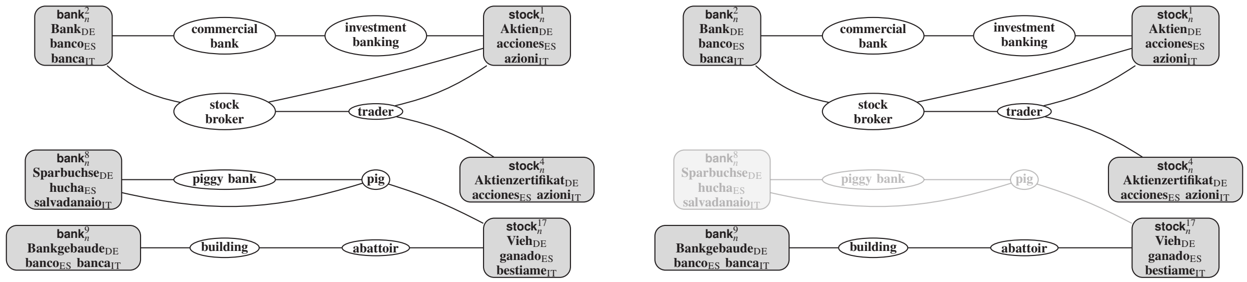
Example. We now describe the execution of our algorithm by means of an example. Suppose we are given as input the English word pair bank-stock. In the first phase of our algorithm (line 1), we build the initial semantic graph from the senses of these input words (Figure 1(a)). Both words are highly polysemous and, accordingly, the semantic graph contains different senses, including less frequent ones such as the sense of bank as 'container for money' and the 'animal' sense of stock. Note that the graph also contains noisy paths resulting from spurious semantic relations (e.g. the edge between piggy bank and pig). Next, we collect the translations for the input words (lines 2–9): these include $\text{banco}_{\mathrm{ES}}$, $\text{banca}_{\mathrm{IT}}$ and $\text{Bank}_{\mathrm{DE}}$ for bank, and $\text{acciones}_{\mathrm{ES}}$, $\text{magazzino}_{\mathrm{IT}}$ and $\text{Lager}_{\mathrm{DE}}$ for stock. We then intersect the initial graph with the graphs created from the translations (lines 10–23): this has the effect of filtering out at successive stages infrequent senses and noisy relations, and produces a graph which contains the correct senses for the input pair, namely the financial ones, as well as 'strong' semantic relations. In particular, the intersection with the Spanish-English word pair $\text{banco}_{\mathrm{ES}}$-$\text{stock}_{\mathrm{EN}}$ (Figure 1(b)) leads to the exclusion of the 'money container' sense of bank and the spurious paths that include it. Next, by means of the intersection with the German-Spanish pair $\text{Bank}_{\mathrm{DE}}$-$\text{acciones}_{\mathrm{ES}}$ (Figure 1(c)), we are also able to remove senses and paths for the 'building' and 'animal' senses of bank and stock, respectively. Note that this could not have been achieved using either WordNet's first sense (as it would have selected the 'river' sense of bank) or by simply selecting the synsets containing the most frequent translations (since it would have yielded an empty graph, due to $\text{magazzino}_{\mathrm{IT}}$ and $\text{Lager}_{\mathrm{DE}}$ covering only the 'inventory' sense of stock). The resulting graph involves the financial senses of the input words and their relations: this is used to compute semantic relatedness in lines 24–30, yielding a final score of 0.33 based on the shortest path bank_n^2 — stock broker — stock_n^1.

Experiments

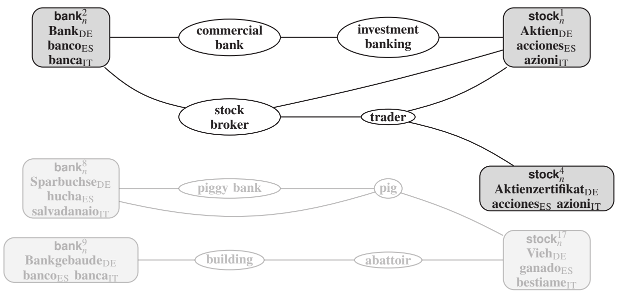
Experimental setting. We benchmark our approach on standard datasets designed to evaluate monolingual and bilingual semantic relatedness. In order to evaluate semantic relatedness across languages we use the data from Hassan and Mihalcea (2009), which consists of the translation of two standard English datasets, namely the Miller and Charles (1991, M&C) and WordSimilarity-353 Test Collection (Finkelstein et al. 2002, WS-353TC) data, into three different languages (Spanish, Romanian, Arabic). Starting with these two English datasets, Hassan and Mihalcea (2009) created their multilingual versions by translating each word pair into the non-English languages – e.g., the word pair bird and cock is translated into Spanish as pajaro and gallo. The cross-lingual data were instead created by taking, for each English word pair, their translations in different languages, and pairing each word in one language with a word in another language. For instance, two word pairs are created for the English-Spanish dataset, namely $\text{bird}_{\mathrm{EN}}$-$\text{gallo}_{\mathrm{ES}}$ and $\text{cock}_{\mathrm{EN}}$-$\text{pajaro}_{\mathrm{ES}}$. In addition, we use the dataset from Gurevych (2005, Gurevych-65), which is a German translation of the Rubenstein and Goodenough (1965) data. To evaluate performance we follow Hassan and Mihalcea (2009) and report both Pearson product-moment (r) and Spearman rank (ρ) correlation coefficients between the relatedness scores and the corresponding human judgments.

(a) Semantic graph for the English pair bank-stock.

(b) Semantic graph after intersection with banco_{ES}-stock_{EN}.



(c) Semantic graph after intersection with Bank_{DE}-$\text{acciones}_{\text{ES}}$.



(d) List of BabelNet senses and their glosses (from WordNet).

bank^2_n	financial institution that accepts deposits and channels the money into lending activities
bank^8_n	a container (usually with a slot in the top) for keeping money at home
bank^9_n	a building in which the business of banking transacted
stock^1_n	the capital raised by a corporation through the issue of shares entitling holders to an ownership interest
stock^4_n	a certificate documenting the shareholder's ownership in the corporation
stock^{17}_n	any animals kept for use or profit

Figure 1: Joint multilingual semantic graph construction for the word pair bank-stock.

Results and discussion. In Tables 1–3 we show the performance of two different methods on the M&C, WS-353TC and Gurevych-65 datasets, namely[3]:

- **BabelRelate simple:** a baseline approach that does not exploit any language other than those of the two input words, i.e., it computes semantic relatedness using only the semantic graph built from the input words (i.e., we skip lines 15–23 of Algorithm 1).
- **BabelRelate joint:** the full approach described in Algorithm 1, which uses the semantic graphs of the input words, as well as their translations in other languages.

We compare our results with state-of-the-art methods from the literature. For the M&C and WS-353TC datasets, we report the original figures from Hassan and Mihalcea (2009), who proposed a multilingual extension of Explicit Semantic Analysis (ESA) which uses the inter-language links from Wikipedia. On the Gurevych-65 dataset, instead, we compare our performance with different knowledge-based methods, including information content based approaches (Resnik 1999; Lin 1998) and a gloss-based method (Lesk 1986), that Gurevych (2005) applied to GermaNet (Lemnitzer and Kunze 2002)[4], as well as an application of taxonomy-based measures to the category graph of Wikipedia (Ponzetto and Strube 2007).

Resource	Pearson's r	Spearman's ρ
GermaNet†	.49–.66	–
Wikipedia (categories)†	.33–.65	–
BabelRelate simple	.63	.66
BabelRelate joint	.79	.83

Table 3: Results on the Gurevych-65 dataset. † indicates results reported in Ponzetto and Strube (2007).

The results are consistent across all datasets and measures. In comparison with other methods from the literature our baseline, which does not use any multilingual information in a joint way, already achieves very good results: on the M&C and WS-353TC data, we are, in fact, able to perform better than Hassan and Mihalcea (2009) on all languages – up to +0.42 Pearson and +0.49 Spearman on the M&C data (both on ES-RO, Table 1), and +0.26 Pearson (EN-RO) and +0.30 Spearman (ES-RO, RO-RO) on the WS-353TC dataset (Table 2) – except when evaluating using Spearman on the WS-353TC data. In general, we take these baseline results to indicate the high quality of the relations found in BabelNet. Thanks to a common resource shared across all its languages (i.e. the semantic network of Babel synsets), BabelNet allows us to achieve comparable results for each of these languages. That is, we do not suffer from an unbalanced performance across languages, as Hassan and Mihal-

[3]We use the following language abbreviations: 'EN' for English, 'ES' for Spanish, 'AR' for Arabic, 'RO' for Romanian.

[4]In Table 3, we report the results on the full dataset of 65 word pairs originally provided in Ponzetto and Strube (2007), instead of comparing with the results on subsets of the dataset, e.g. only word pairs covered by GermaNet (Gurevych 2005), or all of GermaNet, Wikipedia and Wiktionary (Zesch, Müller, and Gurevych 2008).

(a) Pearson's r

	EN	EN	EN	EN	ES	ES	ES	AR	AR	RO
	EN	ES	AR	RO	ES	AR	RO	AR	RO	RO
Hassan and Mihalcea (2009)	.58	.43	.32	.50	.44	.20	.38	.36	.32	.58
BabelRelate simple	.79	.81	.57	.79	.82	.57	.80	.40	.56	.70
BabelRelate joint	**.89**	**.86**	**.74**	**.88**	**.83**	**.75**	**.83**	**.69**	**.73**	**.82**

(b) Spearman's ρ

	EN	EN	EN	EN	ES	ES	ES	AR	AR	RO
	EN	ES	AR	RO	ES	AR	RO	AR	RO	RO
Hassan and Mihalcea (2009)	.75	.56	.27	.55	.64	.17	.32	.33	.21	.61
BabelRelate simple	.87	.85	.62	.82	.83	.59	.81	.48	.55	.74
BabelRelate joint	**.90**	**.87**	**.74**	**.89**	**.85**	**.77**	**.86**	**.71**	**.72**	**.82**

Table 1: Results for cross-lingual relatedness on the M&C dataset (Miller and Charles 1991).

(a) Pearson's r

	EN	EN	EN	EN	ES	ES	ES	AR	AR	RO
	EN	ES	AR	RO	ES	AR	RO	AR	RO	RO
Hassan and Mihalcea (2009)	.55	.32	.31	.29	.45	.32	.28	.28	.25	.30
BabelRelate simple	.58	.57	.43	.55	.58	.42	.53	.32	.42	.51
BabelRelate joint	**.59**	**.59**	**.53**	**.56**	**.60**	**.53**	**.55**	**.51**	**.51**	**.53**

(b) Spearman's ρ

	EN	EN	EN	EN	ES	ES	ES	AR	AR	RO
	EN	ES	AR	RO	ES	AR	RO	AR	RO	RO
Hassan and Mihalcea (2009)	**.71**	.55	.35	.38	.50	.29	.30	.26	.20	.28
BabelRelate simple	.64	.63	.50	.62	.63	.49	.60	.39	.49	.58
BabelRelate joint	.65	**.66**	**.61**	**.63**	**.67**	**.61**	**.63**	**.57**	**.58**	**.59**

Table 2: Results for cross-lingual relatedness on the WS-353TC dataset (Finkelstein et al. 2002).

cea (2009) do, arguably because of the different distributions of inter-language links across wikipedias in resource-rich vs. resource-poor languages. BabelNet, in fact, tackles the problem of lexical translation gaps by means of a machine translation system, thus filling such gaps and achieving high coverage for all languages. Manual inspection of the output on the M&C dataset revealed that the lower results for Arabic (e.g. the 0.42 difference in Pearson correlation between ES-ES and AR-AR in Table 1(a)) were caused by its high polysemy, rather than missing senses or translations – e.g. the Arabic word for string in the M&C dataset also translates as tendon and nerve in BabelNet. This, together with a very simple measure such as the one we apply, that is triggered when at least one path between two senses exist, can often lead to an overestimation of the relatedness score due to a spurious path between infrequent senses.

By using BabelNet with our method to jointly exploit various different languages at the same time, we are able not only to improve the baseline results for all language pairs, datasets and measures, but also to achieve the best performance overall – i.e. up to 0.89 Pearson and 0.90 Spearman on the English M&C data (Table 1). As shown in Figure 1, jointly exploiting multilingual semantic graphs allows us to remove infrequent senses and noisy relations, and thus find the core semantic representation of the word pair shared across all languages. As a result of this our method is able to yield consistent improvements in all evaluation settings, with a bigger effect on a highly polysemous language such as Arabic ($+0.29$ Pearson and $+0.23$ Spearman on the M&C data, compared to the simple approach), thus achieving comparable results across all language pairs. These trends are also supported by the performance figures on the Gurevych-65 dataset shown in Table 3, where our baseline attains a performance comparable to that obtained using other resources such as GermaNet and the category system of Wikipedia, while our joint method again achieves overall best results, performing very close to the estimated upper bound of $r = 0.81$ for this task (Gurevych 2005).

Joint use of languages pays off. Since jointly exploiting multiple languages provides a significant boost in all scenarios, we tested whether this is due to a specific subset of languages (e.g., English), or is, instead, language-independent. In order to do this we evaluated our joint approach on each language pair with different subsets of increasing sizes for the set L of languages used to create the multilingual semantic graphs (these were randomly sampled from the full set of languages covered by BabelNet). We report the re-

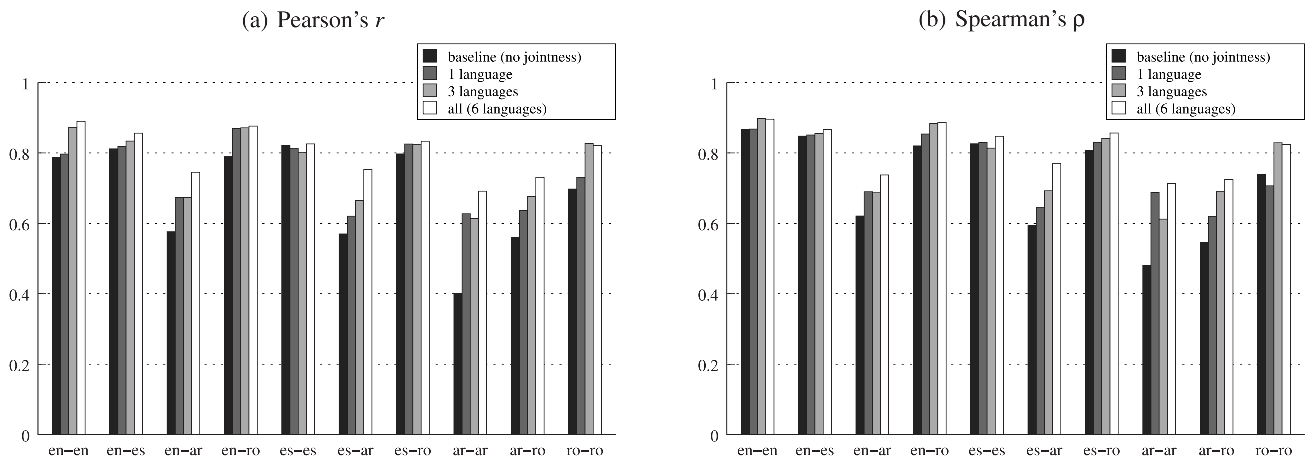


Figure 2: Performance levels for the different subsets of languages used to create the multilingual semantic graphs (M&C data).

sults on the M&C dataset in Figure 2[5]. Similarly to the previous evaluation, the results are consistent across language pairs and measures, and indicate that performance increases along with the number of languages used. That is, *the more languages we use, the higher the performance is*: our results thus not only show that jointness helps when computing semantic relatedness, but also that performance benefits from exploiting ever more languages at the same time. Once again, the effects are more beneficial for highly polysemous languages (cf., e.g., the performance increases for the ES-AR and AR-RO language pairs), since, as we showed in our example, our approach drastically reduces the input words' polysemy by creating a core graph, which is focused on the most frequent translations in all languages.

Related work

Recent years have seen a great deal of work done on computing semantic relatedness. As often happens with many other Natural Language Processing tasks, most of this research was carried out in English, while using WordNet as the *de facto* standard knowledge resource. However, recently, due to the emergence of a whole new spectrum of resources, a variety of methods have been developed to compute semantic relatedness using Wikipedia (Ponzetto and Strube 2007; 2011; Gabrilovich and Markovitch 2009; Milne and Witten 2008; Yeh et al. 2009), Wiktionary (Zesch, Müller, and Gurevych 2008), as well as Web search engines (Chen, Lin, and Wei 2006; Sahami and Heilman 2006; Bollegala, Matsuo, and Ishizuka 2007). Similarly, recent studies have concentrated on evaluating semantic relatedness on languages other than English, such as German (Zesch and Gurevych 2010) and Chinese (Liu and Chen 2010).

The work closest to ours is that of Hassan and Mihalcea (2009), who were the first to introduce the task of cross-lingual semantic relatedness. In contrast to their work, however, what we explore here is the joint contribution obtained by using a multilingual knowledge base for this task, instead of combining a concept vector space model such as Explicit Semantic Analysis (Gabrilovich and Markovitch 2009) with cross-lingual relations harvested from Wikipedia's interlanguage links. Using a wider range of different datasets and language pairs our experiments confirm the seminal findings of Agirre et al. (2009) that knowledge-based approaches to semantic relatedness can compete and even outperform distributional methods in a cross-lingual setting. In addition, in this work we crucially show the beneficial effects of multilingual language jointness for computing semantic relatedness. Our experiments show that joining forces across languages does pay off: we achieve this by exploiting BabelNet, a wide-coverage multilingual lexical knowledge base which is complementary to other resources like WikiNet (Nastase et al. 2010) and MENTA (de Melo and Weikum 2010).

Conclusions

In this paper we presented a knowledge-rich approach to computing semantic relatedness using a multilingual lexical knowledge base. Key to our approach is the exploitation of information from different languages at the same time: we achieve this by means of a graph-based algorithm that combines the semantic graphs of the input words with those of their translations in other languages. The results show not only that information from different languages can help better estimate semantic relatedness across any language pair, thus bridging the performance gap between resource-rich and resource-poor languages, but also that the more languages we use, the better the results we achieve.

As future work we plan to apply our method to a variety of applications which have been shown to benefit from semantic relatedness in a monolingual setting. We believe that a high performance in cross-lingual semantic relatedness will enable a wide range of multilingual applications in Natural Language Processing and Information Retrieval.

[5] We leave out the results on WS-353TC and Gurevych-65 for the sake of brevity. However, they all exhibit the same trend.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Starting Grant MultiJEDI No. 259234.

Thanks also go to Google for access to the University Research Program for Google Translate.

References

Agirre, E.; Alfonseca, E.; Hall, K.; Kravalova, J.; Paşca, M.; and Soroa, A. 2009. A study on similarity and relatedness using distributional and WordNet-based approaches. In *Proc. of NAACL-HLT-09*, 19–27.

Bollegala, D.; Matsuo, Y.; and Ishizuka, M. 2007. Measuring semantic similarity between words using web search engines. In *Proc. of WWW-07*, 757–766.

Botafogo, R. A.; Rivlin, E.; and Shneiderman, B. 1992. Structural analysis of hypertexts: Identifying hierarchies and useful metrics. *ACM Transactions on Information Systems* 10(2):142–180.

Budanitsky, A., and Hirst, G. 2006. Evaluating WordNet-based measures of semantic distance. *Computational Linguistics* 32(1):13–47.

Chen, H.-H.; Lin, M.-S.; and Wei, Y.-C. 2006. Novel association measures using web search with double checking. In *Proc. of COLING-ACL-06*, 1009–1016.

Cimiano, P.; Schultz, A.; Sizov, S.; Sorg, P.; and Staab, S. 2009. Explicit vs. latent concept models for cross-language information retrieval. In *Proc. of IJCAI-09*, 1513–1518.

de Melo, G., and Weikum, G. 2010. MENTA: Inducing multilingual taxonomies from Wikipedia. In *Proc. of CIKM-10*, 1099–1108.

Dumais, S.; Landauer, T.; and Littman, M. 1996. Automatic cross-linguistic information retrieval using Latent Semantic Indexing. In *SIGIR-96 Workshop on Cross-Linguistic Information Retrieval*, 16–23.

Egozi, O.; Markovitch, S.; and Gabrilovich, E. 2011. Concept-based information retrieval using Explicit Semantic Analysis. *ACM Transactions on Information Systems* 29(2):8:1–8:34.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Mass.: MIT Press.

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

Gabrilovich, E., and Markovitch, S. 2009. Wikipedia-based semantic interpretation for natural language processing. *Journal of Artificial Intelligence Research* 34:443–498.

Gurevych, I. 2005. Using the structure of a conceptual network in computing semantic relatedness. In *Proc. of IJCNLP-05*, 767–778.

Hassan, S., and Mihalcea, R. 2009. Cross-lingual semantic relatedness using encyclopedic knowledge. In *Proc. of EMNLP-09*, 1192–1201.

Krötzsch, M.; Vrandečić, D.; Völkel, M.; Haller, H.; and Studer, R. 2007. Semantic Wikipedia. *Journal of Web Semantics* 5(4):251–261.

Leacock, C., and Chodorow, M. 1998. Combining local context and WordNet similarity for word sense identification. In Fellbaum, C., ed., *WordNet. An Electronic Lexical Database*. Cambridge, Mass.: MIT Press. chapter 11, 265–283.

Lemnitzer, L., and Kunze, C. 2002. GermaNet – representation, visualization, application. In *Proc. of LREC '02*.

Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proc. of SIGDOC '86*, 24–26.

Lin, D. 1998. An information-theoretic definition of similarity. In *Proc. of ICML-98*, 296–304.

Liu, H., and Chen, Y. 2010. Computing semantic relatedness between named entities using Wikipedia. In *Proc. of AICI-10*, 388–392.

Miller, G. A., and Charles, W. G. 1991. Contextual correlates of semantic similarity. *Language and Cognitive Processes* 6(1):1–28.

Milne, D., and Witten, I. H. 2008. An effective, low-cost measure of semantic relatedness obtained from Wikipedia links. In *Proceedings of the AAAI-08 WikiAI Workshop*, 25–30.

Nastase, V.; Strube, M.; Börschinger, B.; Zirn, C.; and Elghafari, A. 2010. WikiNet: A very large scale multi-lingual concept network. In *Proc. of LREC '10*.

Navigli, R., and Crisafulli, G. 2010. Inducing word senses to improve web search result clustering. In *Proc. of EMNLP-10*, 116–126.

Navigli, R., and Lapata, M. 2010. An experimental study on graph connectivity for unsupervised Word Sense Disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32(4):678–692.

Navigli, R., and Ponzetto, S. P. 2010. BabelNet: Building a very large multilingual semantic network. In *Proc. of ACL-10*, 216–225.

Ponzetto, S. P., and Strube, M. 2007. Knowledge derived from Wikipedia for computing semantic relatedness. *Journal of Artificial Intelligence Research* 30:181–212.

Ponzetto, S. P., and Strube, M. 2011. Taxonomy induction based on a collaboratively built knowledge repository. *Artificial Intelligence* 175:1737–1756.

Potthast, M.; Stein, B.; and Anderka, M. 2008. A Wikipedia-based multilingual retrieval model. In *Proc. of ECIR-08*, 522–530.

Reisinger, J., and Paşca, M. 2011. Fine-grained class label markup of search queries. In *Proc. of ACL-11*, 1200–1209.

Resnik, P. 1999. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* 11:95–130.

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

Sahami, M., and Heilman, T. D. 2006. A web-based kernel function for measuring the similarity of short text snippets. In *Proc. of WWW-06*, 377–386.

Toderici, G.; Aradhye, H.; Paşca, M.; Sbaiz, L.; and Yagnik, J. 2010. Finding meaning on YouTube: Tag recommendation and category discovery. In *Proc. of CVPR 2010*, 3447–3454.

Wu, Z., and Palmer, M. 1994. Verb semantics and lexical selection. In *Proc. of ACL-94*, 133–138.

Yeh, E.; Ramage, D.; Manning, C. D.; Agirre, E.; and Soroa, A. 2009. WikiWalk: Random walks on Wikipedia for semantic relatedness. In *Proceedings of the ACL-09 TextGraphs-4 Workshop*, 41–49.

Zesch, T., and Gurevych, I. 2010. Wisdom of crowds versus wisdom of linguists – measuring the semantic relatedness of words. *Natural Language Engineering* 16(1):25–59.

Zesch, T.; Müller, C.; and Gurevych, I. 2008. Using Wiktionary for computing semantic relatedness. In *Proc. of AAAI-08*, 861–867.