

UMCC_DLSI: Textual Similarity based on Lexical-Semantic features

**Alexander Chávez, Antonio Fernández Orquín,
Héctor Dávila, Yoan Gutiérrez, Armando
Collazo, José I. Abreu**

DI, University of Matanzas
Autopista a Varadero km 3 ½
Matanzas, Cuba.

{alexander.chavez, tony,
hector.davila, yoan.gutierrez,
armando.collazo, jose.abreu}@umcc.cu

Andrés Montoyo, Rafael Muñoz

DLSI, University of Alicante Carretera de
San Vicente S/N Alicante, Spain.
{montoyo, rafael}@dlsi.ua.es

Abstract

This paper describes the specifications and results of UMCC_DLSI system, which participated in the Semantic Textual Similarity task (STS) of SemEval-2013. Our supervised system uses different types of lexical and semantic features to train a Bagging classifier used to decide the correct option. Related to the different features we can highlight the resource ISR-WN used to extract semantic relations among words and the use of different algorithms to establish semantic and lexical similarities. In order to establish which features are the most appropriate to improve STS results we participated with three runs using different set of features. Our best run reached the position 44 in the official ranking, obtaining a general correlation coefficient of 0.61.

1 Introduction

SemEval-2013 (Agirre *et al.*, 2013) presents the task Semantic Textual Similarity (STS) again. In STS, the participating systems must examine the degree of semantic equivalence between two sentences. The goal of this task is to create a unified framework for the evaluation of semantic textual similarity modules and to characterize their impact on NLP applications.

STS is related to Textual Entailment (TE) and Paraphrase tasks. The main difference is that STS assumes bidirectional graded equivalence between the pair of textual snippets.

In case of TE, the equivalence is directional (e.g. a student is a person, but a person is not

necessarily a student). In addition, STS differs from TE and Paraphrase in that, rather than being a binary yes/no decision, STS is a similarity-graded notion (e.g. a student is more similar to a person than a dog to a person).

This graded bidirectional is useful for NLP tasks such as Machine Translation (MT), Information Extraction (IE), Question Answering (QA), and Summarization. Several semantic tasks could be added as modules in the STS framework, “*such as Word Sense Disambiguation and Induction, Lexical Substitution, Semantic Role Labeling, Multiword Expression detection and handling, Anaphora and Co-reference resolution, Time and Date resolution and Named Entity, among others*”¹

1.1 Description of 2013 pilot task

This edition of SemEval-2013 remain with the same classification approaches that in their first version in 2012. The output of different systems was compared to the reference scores provided by SemEval-2013 gold standard file, which range from five to zero according to the next criterions²: (5) “The two sentences are equivalent, as they mean the same thing”. (4) “The two sentences are mostly equivalent, but some unimportant details differ”. (3) “The two sentences are roughly equivalent, but some important information differs/missing”. (2) “The two sentences are not equivalent, but share some details”. (1) “The two sentences are not

¹ <http://www.cs.york.ac.uk/semeval-2012/task6/>

² <http://www.cs.york.ac.uk/semeval-2012/task6/data/uploads/datasets/train-readme.txt>

equivalent, but are on the same topic”. (0) “The two sentences are on different topics”.

After this introduction, the rest of the paper is organized as follows. Section 3 shows the Related Works. Section 4 presents our system architecture and description of the different runs. In section 4 we describe the different features used in our system. Results and a discussion are provided in Section 5 and finally we conclude in Section 6.

2 Related Works

There are more extensive literature on measuring the similarity between documents than to between sentences. Perhaps the most recently scenario is constituted by the competition of SemEval-2012 task 6: A Pilot on Semantic Textual Similarity (Aguirre and Cerd, 2012). In SemEval-2012, there were used different tools and resources like stop word list, multilingual corpora, dictionaries, acronyms, and tables of paraphrases, “*but WordNet was the most used resource, followed by monolingual corpora and Wikipedia*” (Aguirre and Cerd, 2012).

According to Aguirre, Generic NLP tools were widely used. Among those that stand out were tools for lemmatization and POS-tagging (Aguirre and Cerd, 2012). On a smaller scale word sense disambiguation, semantic role labeling and time and date resolution. In addition, Knowledge-based and distributional methods were highly used. Aguirre and Cerd remarked on (Aguirre and Cerd, 2012) that alignment and/or statistical machine translation software, lexical substitution, string similarity, textual entailment and machine translation evaluation software were used to a lesser extent. It can be noted that machine learning was widely used to combine and tune components.

Most of the knowledge-based methods “*obtain a measure of relatedness by utilizing lexical resources and ontologies such as WordNet (Miller et al., 1990b) to measure definitional overlap, term distance within a graphical taxonomy, or term depth in the taxonomy as a measure of specificity*” (Banea et al., 2012).

Some scholars as in (Corley and Mihalcea, June 2005) have argue “*the fact that a comprehensive metric of text semantic similarity should take into account the relations between*

words, as well as the role played by the various entities involved in the interactions described by each of the two sentences”. This idea is resumed in the Principle of Compositionality, this principle posits that the meaning of a complex expression is determined by the meanings of its constituent expressions and the rules used to combine them (Werning et al., 2005). Corley and Mihalcea in this article combined metrics of word-to-word similarity, and language models into a formula and they pose that this is a potentially good indicator of the semantic similarity of the two input texts sentences. They modeled the semantic similarity of a sentence as a function of the semantic similarity of the component words (Corley and Mihalcea, June 2005).

One of the top scoring systems at SemEval-2012 (Šarić et al., 2012) tended to use most of the aforementioned resources and tools. They predict the human ratings of sentence similarity using a support-vector regression model with multiple features measuring word-overlap similarity and syntax similarity. They also compute the similarity between sentences using the semantic alignment of lemmas. First, they compute the word similarity between all pairs of lemmas from first to second sentence, using either the knowledge-based or the corpus-based semantic similarity. They named this method Greedy Lemma Aligning Overlap.

Daniel Bär presented the UKP system, which performed best in the Semantic Textual Similarity (STS) task at SemEval-2012 in two out of three metrics. It uses a simple log-linear regression model, trained on the training data, to combine multiple text similarity measures of varying complexity.

3 System architecture and description of the runs

As we can see in Figure 1, our three runs begin with the pre-processing of SemEval-2013’s training set. Every sentence pair is tokenized, lemmatized and POS-tagged using Freeling 2.2 tool (Atserias et al., 2006). Afterwards, several methods and algorithms are applied in order to extract all features for our Machine Learning System (MLS). Each run uses a particular group of features.

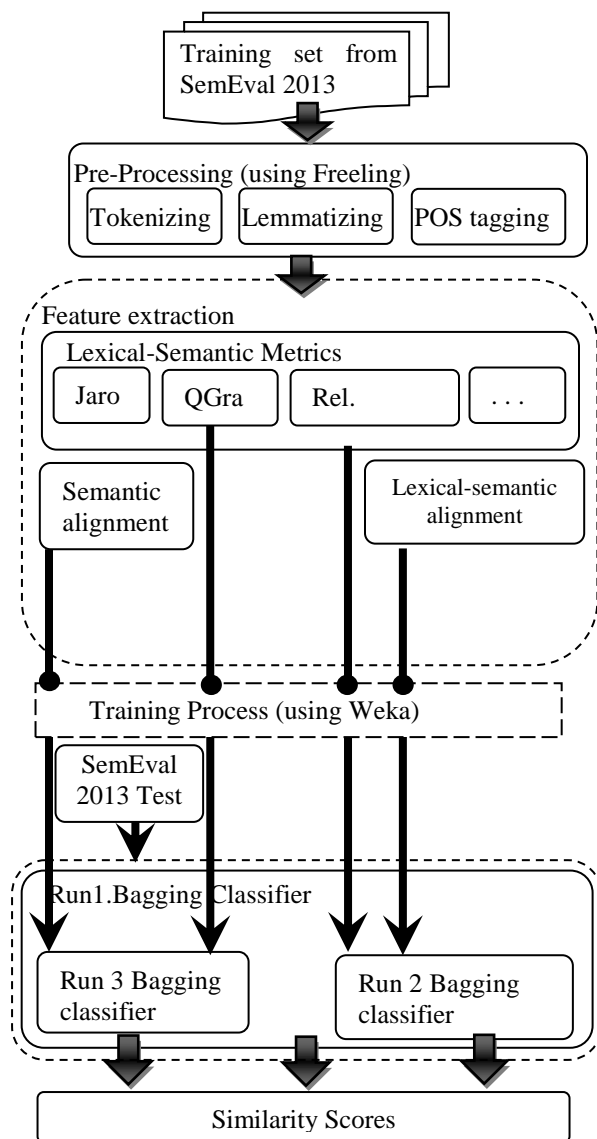


Figure 1. System Architecture.

The Run 1 (named **MultiSemLex**) is our main run. This takes into account all extracted features and trains a model with a Bagging classifier (Breiman, 1996) (using REPTree). The training corpus has been provided by SemEval-2013 competition, in concrete by the Semantic Textual Similarity task.

The Run 2 (named **MultiLex**) and Run 3 (named **MultiSem**) use the same classifier, but including different features. Run 2 uses (see Figure 1) features extracted from Lexical-Semantic Metrics (LS-M) described in section 4.1, and Lexical-Semantic Alignment (LS-A) described in section 4.2.

On the other hand, Run 3 uses features extracted only from Semantic Alignment (SA) described in section 4.3.

As a result, we obtain three trained models capable to estimate the similarity value between two phrases.

Finally, we test our system with the SemEval-2013 test set (see Table 14 with the results of our three runs). The following section describes the features extraction process.

4 Description of the features used in the Machine Learning System

Many times when two phrases are very similar, one sentence is in a high degree lexically overlapped by the other. Inspired in this fact we developed various algorithms, which measure the level of overlapping by computing a quantity of matching words in a pair of phrases. In our system, we used as features for a MLS lexical and semantic similarity measures. Other features were extracted from a lexical-semantic sentences alignment and a variant using only a semantic alignment.

4.1 Similarity measures

We have used well-known string based similarity measures like: Needleman-Wunch (sequence alignment), Smith-Waterman (sequence alignment), Smith-Waterman-Gotoh, Smith-Waterman-Gotoh-Windowed-Affine, Jaro, Jaro-Winkler, Chapman-Length-Deviation, Chapman-Mean-Length, QGram-Distance, Block-Distance, Cosine Similarity, Dice Similarity, Euclidean Distance, Jaccard Similarity, Matching Coefficient, Monge-Elkan and Overlap-Coefficient. These algorithms have been obtained from an API (*Application Program Interface*) SimMetrics library v1.5 for .NET 2.0³. We obtained 17 features for our MLS from these similarity measures.

Using Levenshtein's edit distance (LED), we computed also two different algorithms in order to obtain the alignment of the phrases. In the first one, we considered a value of the alignment as the LED between two sentences. Contrary to (Tatu *et al.*, 2006), we do not remove the punctuation or stop words from the sentences,

³ Copyright (c) 2006 by Chris Parkinson, available in <http://sourceforge.net/projects/simmetrics/>

neither consider different cost for transformation operation, and we used all the operations (deletion, insertion and substitution).

The second one is a variant that we named Double Levenshtein's Edit Distance (DLED) (see Table 9 for detail). For this algorithm, we used LED to measure the distance between the phrases, but in order to compare the words, we used LED again (Fernández *et al.*, 2012; Fernández Orquín *et al.*, 2009).

Another distance we used is an extension of LED named Extended Distance (in spanish distancia extendida (DEx)) (see (Fernández *et al.*, 2012; Fernández Orquín *et al.*, 2009) for details). This algorithm is an extension of the Levenshtein's algorithm, with which penalties are applied by considering what kind of transformation (insertion, deletion, substitution, or non-operation) and the position it was carried out, along with the character involved in the operation. In addition to the cost matrixes used by Levenshtein's algorithm, DEx also obtains the Longest Common Subsequence (LCS) (Hirschberg, 1977) and other helpful attributes for determining similarity between strings in a single iteration. It is worth noting that the inclusion of all these penalizations makes the DEx algorithm a good candidate for our approach.

In our previous work (Fernández Orquín *et al.*, 2009), DEx demonstrated excellent results when it was compared with other distances as (Levenshtein, 1965), (Needleman and Wunsch, 1970), (Winkler, 1999). We also used as a feature the Minimal Semantic Distances (Breadth First Search (BFS)) obtained between the most relevant concepts of both sentences. The relevant concepts pertain to semantic resources ISR-WN (Gutiérrez *et al.*, 2011; 2010a), as WordNet (Miller *et al.*, 1990a), WordNet Affect (Strapparava and Valitutti, 2004), SUMO (Niles and Pease, 2001) and Semantic Classes (Izquierdo *et al.*, 2007). Those concepts were obtained after having applied the Association Ratio (AR) measure between concepts and words over each sentence. (We refer reader to (Gutiérrez *et al.*, 2010b) for a further description).

Another attribute obtained by the system was a value corresponding with the sum of the smaller distances (using QGram-Distance) between the

words or the lemmas of the phrase one with each words of the phrase two.

As part of the attributes extracted by the system, was also the value of the sum of the smaller distances (using Levenshtein) among stems, chunks and entities of both phrases.

4.2 Lexical-Semantic alignment

Another algorithm that we created is the Lexical-Semantic Alignment. In this algorithm, we tried to align the phrases by its lemmas. If the lemmas coincide we look for coincidences among parts-of-speech⁴ (POS), and then the phrase is realigned using both. If the words do not share the same POS, they will not be aligned. To this point, we only have taken into account a lexical alignment. From now on, we are going to apply a semantic variant. After all the process, the non-aligned words will be analyzed taking into account its WordNet's relations (synonymy, hyponymy, hyperonymy, derivationally-related-form, similar-to, verbal group, entailment and cause-to relation); and a set of equivalences like abbreviations of months, countries, capitals, days and currency. In case of hyperonymy and hyponymy relation, words are going to be aligned if there is a word in the first sentence that is in the same relation (hyperonymy or hyponymy) with another one in the second sentence. For the relations "cause-to" and "implication" the words will be aligned if there is a word in the first sentence that causes or implicates another one in the second sentence. All the other types of relations will be carried out in bidirectional way, that is, there is an alignment if a word of the first sentence is a synonymous of another one belonging to the second one or vice versa.

Finally, we obtain a value we called alignment relation. This value is calculated as $FAV = NAW / NWSP$. Where FAV is the final alignment value, NAW is the number of aligned words, and $NWSP$ is the number of words of the shorter phrase. The FAV value is also another feature for our system. Other extracted attributes they are the quantity of aligned words and the quantity of not aligned words. The core of the alignment is carried out in different ways, which

⁴ (noun, verb, adjective, adverbs, prepositions, conjunctions, pronouns, determinants, modifiers, etc.)

are obtained from several attributes. Each way can be compared by:

- the part-of-speech.
- the morphology and the part-of-speech.
- the lemma and the part-of-speech.
- the morphology, part-of-speech, and relationships of WordNet.
- the lemma, part-of-speech, and relationships of WordNet.

4.3 Semantic Alignment

This alignment method depends on calculating the semantic similarity between sentences based on an analysis of the relations, in ISR-WN, of the words that fix them.

First, the two sentences are pre-processed with Freeling and the words are classified according to their POS, creating different groups.

The distance between two words will be the distance, based on WordNet, of the most probable sense of each word in the pair, on the contrary of our previously system in SemEval 2012. In that version, we assumed the selected sense after apply a double Hungarian Algorithm (Kuhn, 1955), for more details please refer to (Fernández *et al.*, 2012). The distance is computed according to the equation (1):

$$d(x, y) = \sum_{i=0}^{m} w * r(L[i], L[i + 1]); \quad (1)$$

Where L is the collection of synsets corresponding to the minimum path between nodes x and y , m is the length of L subtracting one, r is a function that search the relation connecting x and y nodes, w is a weight associated to the relation searched by r (see Table 1).

Relation	Weight
Hyponym, Hypernym	2
Member_Holonym, Member_Meronym, Cause, Entailment	5
Similar_To	10
Antonym	200
Other relation different to Synonymy	60

Table 1. Weights applied to WordNet relations.

Table 1 shows the weights associated to WordNet relations between two synsets.

Let us see the following example:

- We could take the pair 99 of corpus MSRvid (from training set of SemEval-2013) with a littler transformation in

order to a better explanation of our method.

Original pair

A: A polar bear is running towards a group of walruses.

B: A polar bear is chasing a group of walruses.

Transformed pair:

A₁: A polar bear runs towards a group of cats.

B₁: A wale chases a group of dogs.

Later on, using equation (1), a matrix with the distances between all groups of both phrases is created (see Table 2).

GROUPS	polar	bear	runs	towards	group	cats
wale	Dist:=3	Dist:=2	Dist:=3	Dist:=5		Dist:=2
chases	Dist:=4	Dist:=3	Dist:=2	Dist:=4		Dist:=3
group					Dist:=0	
dogs	Dist:=3	Dist:=1	Dist:=4	Dist:=4		Dist:=1

Table 2. Distances between groups.

Using the Hungarian Algorithm (Kuhn, 1955) for Minimum Cost Assignment, each group of the first sentence is checked with each element of the second sentence, and the rest is marked as words that were not aligned.

In the previous example the words “toward” and “polar” are the words that were not aligned, so the number of non-aligned words is two. There is only one perfect match: “group-group” (match with cost=0). The length of the shortest sentence is four. The Table 3 shows the results of this analysis.

Number of exact coincidence	Total Distances of optimal Matching	Number of non-aligned Words
1	5	2

Table 3. Features from the analyzed sentences.

This process has to be repeated for nouns (see Table 4), verbs, adjective, adverbs, prepositions, conjunctions, pronouns, determinants, modifiers, digits and date times. On the contrary, the tables have to be created only with the similar groups of the sentences. Table 4 shows features extracted from the analysis of nouns.

GROUPS	bear	group	cats
wale	Dist := 2		Dist := 2
group		Dist := 0	
dogs	Dist := 1		Dist := 1

Table 4. Distances between groups of nouns.

Number of exact coincidence	Total Distances of optimal Matching	Number of non-aligned Words
1	3	0

Table 5. Feature extracted from analysis of nouns.

Several attributes are extracted from the pair of sentences (see Table 3 and Table 5). Three attributes considering only verbs, only nouns, only adjectives, only adverbs, only prepositions, only conjunctions, only pronouns, only determinants, only modifiers, only digits, and only date times. These attributes are:

- Number of exact coincidences
- Total distance of matching
- Number of words that do not match

Many groups have particular features according to their parts-of-speech. The group of the nouns has one more feature that indicates if the two phrases have the same number (plural or singular). For this feature, we take the average of the number of each noun in the phrase like a number of the phrase.

For the group of adjectives we added a feature indicating the distance between the nouns that modify it from the aligned adjectives, respectively.

For the verbs, we search the nouns that precede it, and the nouns that are next of the verb, and we define two groups. We calculated the distance to align each group with every pair of aligned verbs. The verbs have other feature that specifies if all verbs are in the same verbal time.

With the adverbs, we search the verb that is modified by it, and we calculate their distance from all alignment pairs.

With the determinants and the adverbs we detect if any of the alignment pairs are expressing negations (like don't, or do not) in both cases or not. Finally, we determine if the two phrases have the same principal action. For all this new features, we aid with Freeling tool.

As a result, we finally obtain 42 attributes from this alignment method. It is important to remark that this alignment process searches to solve, for each word from the rows (see Table 4) it has a respectively word from the columns.

4.4 Description of the alignment feature

From the alignment process, we extract different features that help us a better result of our MLS. Table 6 shows the group of features with lexical

and semantic support, based on WordNet relation (named F1). Each of them were named with a prefix, a hyphen and a suffix. Table 7 describes the meaning of every prefix, and Table 8 shows the meaning of the suffixes.

Features
CPA_FCG, CPNA_FCG, SIM_FCG, CPA_LCG, CPNA_LCG, SIM_LCG, CPA_FCGR, CPNA_FCGR, SIM_FCGR, CPA_LCGR, CPNA_LCGR, SIM_LCGR

Table 6. F1. Semantic feature group.

Prefixes	Descriptions
CPA	Number of aligned words.
CPNA	Number of non-aligned words.
SIM	Similarity

Table 7. Meaning of each prefixes.

Prefixes	Compared words for...
FCG	Morphology and POS
LCG	Lemma and POS
FCGR	Morphology, POS and WordNet relation.
LCGR	Lemma, POS and WordNet relation.

Table 8. Suffixes for describe each type of alignment.

Features	Descriptions
LevForma	Levenshtein Distance between two phrases comparing words by morphology
LevLema	The same as above, but now comparing by lemma.
LevDoble	Idem, but comparing again by Levenshtein and accepting words match if the distance is ≤ 2 .
DEx	Extended Distance
NormLevF, NormLevL	Normalized forms of LevForma and LevLema.

Table 9. F2. Lexical alignment measures.

Features
NWunch, SWaterman, SWGotoh, SWGAffine, Jaro, JaroW, CLDeviation, CMLength, QGramD, BlockD, CosineS, DiceS, EuclideanD, JaccardS, MaCoef, MongeElkan, OverlapCoef.

Table 10. Lexical Measure from SimMetrics library.

Features	Descriptions
AxAQGD_L	All against all applying QGramD and comparing by lemmas of the words.
AxAQGD_F	Same as above, but applying QGramD and comparing by morphology.
AxAQGD_LF	Idem, not only comparing by lemma but also by morphology.
AxAlev_LF	All against all applying Levenhstein

	comparing by morphology and lemmas.
AxA_Stems	Idem, but applying Levenhstein comparing by the stems of the words.

Table 11. Aligning all against all.

Other features we extracted were obtained from the following similarity measures (named F2) (see Table 9 for detail).

We used another group named F3, with lexical measure extracted from SimMetric library (see Table 10 for detail).

Finally we used a group of five feature (named F4), extracted from all against all alignment (see Table 11 for detail).

4.5 Description of the training phase

For the training process, we used a supervised learning framework, including all the training set as a training corpus. Using ten-fold cross validation with the classifier mentioned in section 3 (experimentally selected).

As we can see in Table 12, the attributes corresponding with the Test 1 (only lexical attributes) obtain 0.7534 of correlation. On the other side, the attributes of the Test 2 (lexical features with semantic support) obtain 0.7549 of correlation, and all features obtain 0.7987. Being demonstrated the necessity to tackle the problem of the similarity from a multidimensional point of view (see Test 3 in the Table 12).

Features	Correlation on the training data of SemEval-2013		
	Test 1	Test 2	Test 3
F1		0.7549	0.7987
F2			
F3	0.7534		
F4			

Table 12. Features influence. Gray cells mean features are not taking into account.

5 Result and discussion

Semantic Textual Similarity task of SemEval-2013 offered two official measures to rank the systems⁵: Mean- the main evaluation value, Rank- gives the rank of the submission as ordered by the "mean" result.

⁵http://ixa2.si.ehu.es/sts/index.php?option=com_content&view=article&id=53&Itemid=61

Test data for the core test datasets, coming from the following:

Corpus	Description
Headlines:	news headlines mined from several news sources by European Media Monitor using the RSS feed.
OnWN:	mapping of lexical resources OnWN. The sentences are sense definitions from WordNet and OntoNotes.
FNWN:	the sentences are sense definitions from WordNet and FrameNet.
SMT:	SMT dataset comes from DARPA GALE HTER and HyTER. One sentence is a MT output and the other is a reference translation where a reference is generated based on human post editing.

Table 13. Test Core Datasets.

Using these measures, our second run (Run 2) obtained the best results (see Table 14). As we can see in Table 14, our lexical run has obtained our best result, given at the same time worth result in our other runs. This demonstrates that tackling this problem with combining multiple lexical similarity measure produce better results in concordance to this specific test corpora.

To explain Table 14 we present following descriptions: caption in top row mean: 1- Headlines, 2- OnWN, 3- FNWN, 4- SMT and 5- mean.

Run	1	R	2	R	3	R	4	R	5	R
1	0.5841	60	0.4847	54	0.2917	52	0.2855	66	0.4352	58
2	0.6168	55	0.5557	39	0.3045	50	0.3407	28	0.4833	44
3	0.3846	85	0.1342	88	-0.0065	85	0.2736	72	0.2523	87

Table 14. Official SemEval-2013 results over test datasets. Ranking (R).

The Run 1 is our main run, which contains the junction of all attributes (lexical and semantic attributes). Table 14 shows the results of all the runs for a different corpus from test phase. As we can see, Run 1 did not obtain the best results among our runs.

Otherwise, Run 3 uses more semantic analysis than Run 2, from this; Run 3 should get better results than reached over the corpus of FNWN, because this corpus is extracted from FrameNet corpus (Baker *et al.*, 1998) (a semantic network). FNWN provides examples with high semantic content than lexical.

Run 3 obtained a correlation coefficient of 0.8137 for all training corpus of SemEval 2013,

while Run 2 and Run 1 obtained 0.7976 and 0.8345 respectively with the same classifier (Bagging using REPTree, and cross validation with ten-folds). These results present a contradiction between test and train evaluation. We think it is consequence of some obstacles present in test corpora, for example:

In headlines corpus there are great quantity of entities, acronyms and gentilics that we not take into account in our system.

The corpus FNWN presents a non-balance according to the length of the phrases.

In OnWN -test corpus-, we believe that some evaluations are not adequate in correspondence with the training corpus. For example, in line 7 the goal proposed was 0.6, however both phrases are semantically similar. The phrases are:

- the act of lifting something
- the act of climbing something.

We think that 0.6 are not a correct evaluation for this example. Our system result, for this particular case, was 4.794 for Run 3, and 3.814 for Run 2, finally 3.695 for Run 1.

6 Conclusion and future works

This paper have introduced a new framework for recognizing Semantic Textual Similarity, which depends on the extraction of several features that can be inferred from a conventional interpretation of a text.

As mentioned in section 3 we have conducted three different runs, these runs only differ in the type of attributes used. We can see in Table 14 that all runs obtained encouraging results. Our best run was situated at 44th position of 90 runs of the ranking of SemEval-2013. Table 12 and Table 14 show the reached positions for the three different runs and the ranking according to the rest of the teams.

In our participation, we used a MLS that works with features extracted from five different strategies: String Based Similarity Measures, Semantic Similarity Measures, Lexical-Semantic Alignment and Semantic Alignment.

We have conducted the semantic features extraction in a multidimensional context using the resource ISR-WN, the one that allowed us to navigate across several semantic resources (WordNet, WordNet Domains, WordNet Affect, SUMO, SentiWordNet and Semantic Classes).

Finally, we can conclude that our system performs quite well. In our current work, we show that this approach can be used to correctly classify several examples from the STS task of SemEval-2013. Compared with the best run of the ranking (UMBC_EBIQUITY- ParingWords) (see Table 15) our main run has very close results in headlines (1), and SMT (4) core test datasets.

Run	1	2	3	4	5	6
(First)	0.7642	0.7529	0.5818	0.3804	0.6181	1
(Our) RUN 2	0.6168	0.5557	0.3045	0.3407	0.4833	44

Table 15. Comparison with best run (SemEval 2013).

As future work we are planning to enrich our semantic alignment method with Extended WordNet (Moldovan and Rus, 2001), we think that with this improvement we can increase the results obtained with texts like those in OnWN test set.

6.1 Team Collaboration

Is important to remark that our team has been working up in collaboration with INAOE (Instituto Nacional de Astrofísica, Óptica y Electrónica) and LIPN (Laboratoire d'Informatique de Paris-Nord), Université Paris 13 universities, in order to encourage the knowledge interchange and open shared technology. Supporting this collaboration, INAOE-UPV (Instituto Nacional de Astrofísica, Óptica y Electrónica and Universitat Politècnica de València) team, in concrete in INAOE-UPV-run 3 has used our semantic distances for nouns, adjectives, verbs and adverbs, as well as lexical attributes like LevDoble, NormLevF, NormLevL and Ext (see influence of these attributes in Table 12).

Acknowledgments

This research work has been partially funded by the Spanish Government through the project TEXT-MESS 2.0 (TIN2009-13391-C04), "Análisis de Tendencias Mediante Técnicas de Opinión Semántica" (TIN2012-38536-C03-03) and "Técnicas de Deconstrucción en la Tecnologías del Lenguaje Humano" (TIN2012-31224); and by the Valencian Government through the project PROMETEO (PROMETEO/2009/199).

Reference

- Agirre, E.; D. Cer; M. Diab and W. Guo. *SEM 2013 Shared Task: Semantic Textual Similarity including a Pilot on Typed-Similarity. *SEM 2013: The Second Joint Conference on Lexical and Computational Semantics, Association for Computational Linguistics, 2013.
- Aguirre, E. and D. Cerd. SemEval 2012 Task 6:A Pilot on Semantic Textual Similarity. First Joint Conference on Lexical and Computational Semantic (*SEM), Montréal, Canada, Association for Computational Linguistics., 2012. 385-393 p.
- Atserias, J.; B. Casas; E. Comelles; M. González; L. Padró and M. Padró. FreeLing 1.3: Syntactic and semantic services in an opensource NLP library. Proceedings of LREC'06, Genoa, Italy, 2006.
- Baker, C. F.; C. J. Fillmore and J. B. Lowe. The berkeley framenet project. Proceedings of the 17th international conference on Computational linguistics-Volume 1, Association for Computational Linguistics, 1998. 86-90 p.
- Banea, C.; S. Hassan; M. Mohler and R. Mihalcea. UNT:A Supervised Synergistic Approach to SemanticText Similarity. First Joint Conference on Lexical and Computational Semantics (*SEM), Montréal. Canada, Association for Computational Linguistics, 2012. 635-642 p.
- Breiman, L. Bagging predictors Machine learning, 1996, 24(2): 123-140.
- Corley, C. and R. Mihalcea. Measuring the Semantic Similarity of Texts, Association for Computational Linguistic. Proceedings of the ACL Work shop on Empirical Modeling of Semantic Equivalence and Entailment, pages 13-18, June 2005.
- Fernández, A.; Y. Gutiérrez; H. Dávila; A. Chávez; A. González; R. Estrada; Y. Castañeda; S. Vázquez; A. Montoyo and R. Muñoz. UMCC_DLSI: Multidimensional Lexical-Semantic Textual Similarity. {*SEM 2012}: The First Joint Conference on Lexical and Computational Semantics -- Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation {(SemEval 2012)}, Montreal, Canada, Association for Computational Linguistics, 2012. 608-616 p.
- Fernández Orquín, A. C.; J. Díaz Blanco; A. Fundora Rolo and R. Muñoz Guillena. Un algoritmo para la extracción de características lexicográficas en la comparación de palabras. IV Convención Científica Internacional CIUM, Matanzas, Cuba, 2009.
- Gutiérrez, Y.; A. Fernández; A. Montoyo and S. Vázquez. Integration of semantic resources based on WordNet. XXVI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, Universidad Politécnica de Valencia, Valencia, SEPLN 2010, 2010a. 161-168 p. 1135-5948.
- Gutiérrez, Y.; A. Fernández; A. Montoyo and S. Vázquez. UMCC-DLSI: Integrative resource for disambiguation task. Proceedings of the 5th International Workshop on Semantic Evaluation, Uppsala, Sweden, Association for Computational Linguistics, 2010b. 427-432 p.
- Gutiérrez, Y.; A. Fernández; A. Montoyo and S. Vázquez Enriching the Integration of Semantic Resources based on WordNet Procesamiento del Lenguaje Natural, 2011, 47: 249-257.
- Hirschberg, D. S. Algorithms for the longest common subsequence problem J. ACM, 1977, 24: 664-675.
- Izquierdo, R.; A. Suárez and G. Rigau A Proposal of Automatic Selection of Coarse-grained Semantic Classes for WSD Procesamiento del Lenguaje Natural, 2007, 39: 189-196.
- Kuhn, H. W. The Hungarian Method for the assignment problem Naval Research Logistics Quarterly, 1955, 2: 83-97.
- Levenshtein, V. I. Binary codes capable of correcting spurious insertions and deletions of ones. Problems of information Transmission. 1965. pp. 8-17 p.
- Miller, G. A.; R. Beckwith; C. Fellbaum; D. Gross and K. Miller. Five papers on WordNet. Princeton University, Cognositive Science Laboratory, 1990a.
- Miller, G. A.; R. Beckwith; C. Fellbaum; D. Gross and K. Miller Introduction to WordNet: An On-line Lexical Database International Journal of Lexicography, 3(4):235-244., 1990b.
- Moldovan, D. I. and V. Rus Explaining Answers with Extended WordNet ACL, 2001.
- Needleman, S. and C. Wunsch A general method applicable to the search for similarities in the amino acid sequence of two proteins Mol. Biol, 1970, 48(443): 453.
- Niles, I. and A. Pease. Origins of the IEEE Standard Upper Ontology. Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology, Seattle, Washington, USA., 2001.

- Šarić, F.; G. Glavaš; Mladenkaran; J. Šnajder and B. D. Basić. TakeLab: Systems for Measuring Semantic Text Similarity. Montréal, Canada, First Joint Conference on Lexical and Computational Semantics (*SEM), pages 385-393. Association for Computational Linguistics., 2012.
- Strapparava, C. and A. Valitutti. WordNet-Affect: an affective extension of WordNet. Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon, 2004. 1083-1086 p.
- Tatu, M.; B. Iles; J. Slavick; N. Adrian and D. Moldovan. COGEX at the Second Recognizing Textual Entailment Challenge. Proceedings of the Second PASCAL Recognising Textual Entailment Challenge Workshop, Venice, Italy, 2006. 104-109 p.
- Werning, M.; E. Machery and G. Schurz. The Compositionality of Meaning and Content, Volume 1: Foundational issues. ontos verlag [Distributed in] North and South America by Transaction Books, 2005. p. Linguistics & philosophy, Bd. 1. 3-937202-52-8.
- Winkler, W. The state of record linkage and current research problems. Technical Report, Statistical Research Division, U.S, Census Bureau, 1999.