BUAP: Evaluating Compositional Distributional Semantic Models on Full Sentences through Semantic Relatedness and Textual Entailment

Saúl León, Darnes Vilariño, David Pinto, Mireya Tovar, Beatriz Beltrán
Benemérita Universidad Autónoma de Puebla
Faculty of Computer Science
14 Sur y Av. San Claudio, CU
Puebla, Puebla, México
{saul.leon,darnes,dpinto,mtovar,bbeltran}@cs.buap.mx

Abstract

The results obtained by the BUAP team at Task 1 of SemEval 2014 are presented in this paper. The run submitted is a supervised version based on two classification models: 1) We used logistic regression for determining the semantic relatedness between a pair of sentences, and 2) We employed support vector machines for identifying textual entailment degree between the two sentences. The behaviour for the second subtask (textual entailment) obtained much better performance than the one evaluated at the first subtask (relatedness), ranking our approach in the 7th position of 18 teams that participated at the competition.

1 Introduction

The Compositional Distributional Semantic Models (CDSM) applied to sentences aim to approximate the meaning of those sentences with vectors summarizing their patterns of co-occurrence in corpora. In the Task 1 of SemEval 2014, the organizers aimed to evaluate the performance of this kind of models through the following two tasks: semantic relatedness and textual entailment. Semantic relatedness captures the degree of semantic similarity, in this case, between a pair of sentences, whereas textual entailment allows to determine the entailment relation holding between two sentences.

This document is a description paper, therefore, we focus the rest of it on the features and models we used for carrying out the experiments. A complete description of the task and the dataset used are given in Marelli et al. (2014a) and in Marelli et al. (2014b), respectively.

The remaining of this paper is structured as follows. In Section 2 we describe the general model we used for comparing two sentences and the set of the features used for constructing the vectorial representation for each sentence. Section 3 shows how we integrate the features calculated in a single vector which fed a supervised classifier aiming to construct a classication model that solves the two aforementioned problems: semantic relatedness and textual entailment. In the same section we show the obtained results. Finally, in Section 4 we present our findings.

2 Description of the Distributional Semantic Model Used

Given a sentence $S = w_1 w_2 \cdots w_{|S|}$, with w_i a sentence word, we have calculated different correlated terms ($t_{i,j}$) or a numeric vector (V_i) for each word w_i as follows:

1. $\{t_{i,j} | relation(t_{i,j}, w_i)\}$ such as "relation" is one the following dependency relations: "object", "subject" or "property".

2. $\{t_{i,j} | t_{i,j} = c_k \cdots c_{k+n}\}$ with $n = 2, \cdots, 5$, and $c_k \in w_i$; these tokens are also known as n-grams of length n.

3. $\{t_{i,j} | t_{i,j} = c_k \cdots c_{k+((n-1)*r)}\}$ with $n =$

Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014), pages 145–148,
Dublin, Ireland, August 23-24, 2014.

$2, \cdots, 5$, $r = 2, \cdots, 5$, and $c_k \in w_i$; these tokens are also known as $skip$-grams of length n.

4. V_i is obtained by applying the Latent Semantic Analysis (LSA) algorithm implemented in the R software environment for statistical computing and graphics. V_i is basically a vector of values that represent relation of the word w_i with it context, calculated by using a corpus constructed by us, by integrating information from Europarl, Project-Gutenberg and Open Office Thesaurus.

3 A Classification Model for Semantic Relatedness and Textual Entailment based on DSM

Once each sentence has been represented by means of a vectorial representation of patterns, we constructed a single vector, $\overrightarrow{u}$, for each pair of sentences with the aim of capturing the semantic relatedness on the basis of a training corpus.

The entries of this representation vector are calculated by obtaining the semantic similarity between each pair of sentences, using each of the DSM shown in the previous section. In order to calculate each entry, we have found the maximum similarity between each word of the first sentence with respect to the second sentence and, thereafter, we have added all these values, thus, $\overrightarrow{u} = \{f_1, \cdots, f_9\}$.

Given a pair of sentences $S_1 = w_{1,1}w_{2,1} \cdots w_{|S_1|,1}$ and $S_2 = w_{1,2}w_{2,2} \cdots w_{|S_2|,2}$, such as each $w_{i,k}$ is represented according to the correlated terms or numeric vectors established at Section 2, the entry f_i of $\overrightarrow{u}$ is calculated as: $f_l = \sum_{i=1}^{|S_1|} max\{sim(w_{i,1}, w_{j,2})\}$, with $j = 1, \cdots, |S_2|$.

The specific similarity measure ($sim()$) and the correlated term or numeric vector used for each f_l is described as follows:

1. f_1 : $w_{i,k}$ is the "object" of w_i (as defined in 2), and $sim()$ is the maximum similarity obtained by using the following six WordNet similarity metrics offered by NLTK: Leacock & Chodorow (Leacock and Chodorow, 1998), Lesk (Lesk, 1986), Wu & Palmer (Wu and Palmer, 1994), Resnik (Resnik, 1995), Lin (Lin, 1998), and Jiang & Conrath[1] (Jiang and Conrath, 1997).

2. f_2 : $w_{i,k}$ is the "subject" of w_i, and $sim()$ is the maximum similarity obtained by using the same six WordNet similarity metrics.

3. f_3 : $w_{i,k}$ is the "property" of w_i, and $sim()$ is the maximum similarity obtained by using the same six WordNet similarity metrics.

4. f_4 : $w_{i,k}$ is an n-gram containing w_i, and $sim()$ is the cosine similarity measure.

5. f_5 : $w_{i,k}$ is an $skip$-gram containing w_i, and $sim()$ is the cosine similarity measure.

6. f_6 : $w_{i,k}$ is numeric vector obtained with LSA, and $sim()$ is the Rada Mihalcea semantic similarity measure (Mihalcea et al., 2006).

7. f_7 : $w_{i,k}$ is numeric vector obtained with LSA, and $sim()$ is the cosine similarity measure.

8. f_8 : $w_{i,k}$ is numeric vector obtained with LSA, and $sim()$ is the euclidean distance.

9. f_9 : $w_{i,k}$ is numeric vector obtained with LSA, and $sim()$ is the Chebyshev distance.

All these 9 features were introduced to a logistic regression classifier in order to obtain a classification model which allows us to determine the value of relatedness between a new pair of sentences[2]. Here, we use as supervised class, the value of relatedness given to each pair of sentences on the training corpus.

The obtained results for the relatedness subtask are given in Table 1. In columns 2, 3 and 5, a large value signals a more efficient system, but a large MSE (column 4) means a less efficient system. As can be seen, our run obtained the rank 12 of 17, with values slightly below the overall average.

3.1 Textual Entailment

In order to calculate the textual entailment judgment, we have enriched the vectorial representation previously mentioned with synonyms, antonyms and cue-

[1]Natural Language Toolkit of Python; http://www.nltk.org/

[2]We have employed the Weka tool with the default settings for this purpose

Table 1: Results obtained at the substask "Relatedness" of the Semeval 2014 Task 1

TEAM ID	PEARSON	SPEARMAN	MSE	Rank
ECNU_run1	0.82795	0.76892	0.32504	1
StanfordNLP_run5	0.82723	0.75594	0.32300	2
The_Meaning_Factory_run1	0.82680	0.77219	0.32237	3
UNAL-NLP_run1	0.80432	0.74582	0.35933	4
Illinois-LH_run1	0.79925	0.75378	0.36915	5
CECL_ALL_run1	0.78044	0.73166	0.39819	6
SemantiKLUE_run1	0.78019	0.73598	0.40347	7
CNGL_run1	0.76391	0.68769	0.42906	8
UTexas_run1	0.71455	0.67444	0.49900	9
UoW_run1	0.71116	0.67870	0.51137	10
FBK-TR_run3	0.70892	0.64430	0.59135	11
BUAP_run1	**0.69698**	**0.64524**	**0.52774**	**12**
UANLPCourse_run2	0.69327	0.60269	0.54225	13
UQeResearch_run1	0.64185	0.62565	0.82252	14
ASAP_run1	0.62780	0.59709	0.66208	15
Yamraj_run1	0.53471	0.53561	2.66520	16
asjai_run5	0.47952	0.46128	1.10372	17
overall average	0.71876	0.67159	0.63852	8-9
Our difference against the overall average	-2%	-3%	11%	-

words ("no", "not", "nobody" and "none") for detecting negation at the sentences[3]. Thus, if some of these new features exist on the training pair of sentences, we add a boolean value of 1, otherwise we set the feature to zero.

This new set of vectors is introduced to a support vector machine classifier[4], using as class the textual entailment judgment given on the training corpus.

The obtained results for the textual entailment subtask are given in Table 2. Our run obtained the rank 7 of 18, with values above the overall average. We consider that this improvement over the relatedness task was a result of using other features that are quite important for semantic relatedness, such as lexical relations (synonyms and antonyms), and the consideration of the negation phenomenon in the sentences.

4 Conclusions

This paper describes the use of compositional distributional semantic models for solving the problems of semantic relatedness and textual entailment. We proposed different features and measures for that purpose. The obtained results show a competitive approach that may be further improved by considering more lexical relations or other type of semantic similarity measures.

In general, we obtained the 7th place in the official ranking list from a total of 18 teams that participated at the textual entailment subtask. The result at the semantic relatedness subtask could be improved if we were considered to add the new features taken into consideration at the textual entailment subtask, an idea that we will implement in the future.

[3]Synonyms were extracted from WordNet, whereas the antonyms were collected from Wikipedia.

[4]Again, we have employed the weka tool with the default settings for this purpose.

References

Jay J. Jiang and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc of 10th International Conference on Research in Computational Linguistics, ROCLING'97*, pages 19–33, 1997.

Claudia Leacock and Martin Chodorow. Combining local context and wordnet similarity for word sense identification. In Christiane Fellfaum, editor, *MIT Press*, pages 265–283, 1998.

Table 2: Results obtained at the substask “Textual Entailment” of the Semeval 2014 Task 1

TEAM ID	**ACCURACY**	**Rank**
Illinois-LH_run1	84.575	1
ECNU_run1	83.641	2
UNAL-NLP_run1	83.053	3
SemantiKLUE_run1	82.322	4
The_Meaning_Factory_run1	81.591	5
CECL_ALL_run1	79.988	6
BUAP_run1	**79.663**	**7**
UoW_run1	78.526	8
CDT_run1	77.106	9
UIO-Lien_run1	77.004	10
FBK-TR_run3	75.401	11
StanfordNLP_run5	74.488	12
UTexas_run1	73.229	13
Yamraj_run1	70.753	14
asjai_run5	69.758	15
haLF_run2	69.413	16
CNGL_run1	67.201	17
UANLPCourse_run2	48.731	18
Overall average	75.358	11-12
Our difference against the overall average	4.31%	-

Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual International Conference on Systems Documentation*, pages 24–26. ACM, 1986.

Dekang Lin. An information-theoretic definition of similarity. In *Proceedings of the Fifteenth International Conference on Machine Learning*, ICML ’98, pages 296–304, San Francisco, CA, USA, 1998. Morgan Kaufmann Publishers Inc.

Marco Marelli, Luisa Bentivogli, Marco Baroni, Raffaella Bernardi, Stefano Menini, and Roberto Zamparelli. Semeval-2014 task 1: Evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval-2014)*, Dublin, Ireland, 2014a.

Marco Marelli, Stefano Menini, Marco Baroni, Luisa Bentivogli, Raffaella Bernardi, and Roberto Zamparelli. A sick cure for the evaluation of compositional distributional semantic models. In *Proceedings of LREC 2014*, Reykjavik, Iceland, 2014b.

Rada Mihalcea, Courtney Corley, and Carlo Strapparava. Corpus-based and knowledge-based measures of text semantic similarity. In *Proceedings of the 21st National Conference on Artificial Intelligence*, pages 775–780, 2006.

Philip Resnik. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, IJCAI’95, pages 448–453, San Francisco, CA, USA, 1995.

Zhibiao Wu and Martha Stone Palmer. Verb semantics and lexical selection. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 133–138, 1994.