information

MDPI

Article

Modeling the Paraphrase Detection Task over a Heterogeneous Graph Network with Data Augmentation †

Rafael T. Anchiêta [1,2,*], Rogério F. de Sousa [1,2] and Thiago A. S. Pardo [1]

[1] Interinstitutional Center for Computational Linguistics, University of São Paulo, São Carlos 13566-590, Brazil; rogerio.sousa@ifpi.edu.br (R.F.d.S.); taspardo@icmc.usp.br (T.A.S.P.)
[2] Federal Institute of Piauí, Picos 64600-000, Brazil
* Correspondence: rta@ifpi.edu.br

† This paper is an extended version of our paper published in 14th International Conference on the Computational Processing of Portuguese (PROPOR 2020), Evora, Portugal, 2–4 March 2020.

Received: 29 July 2020; Accepted: 28 August 2020; Published: 1 September 2020

Abstract: Paraphrase detection is a Natural-Language Processing (NLP) task that aims at automatically identifying whether two sentences convey the same meaning (even with different words). For the Portuguese language, most of the works model this task as a machine-learning solution, extracting features and training a classifier. In this paper, following a different line, we explore a graph structure representation and model the paraphrase identification task over a heterogeneous network. We also adopt a back-translation strategy for data augmentation to balance the dataset we use. Our approach, although simple, outperforms the best results reported for the paraphrase detection task in Portuguese, showing that graph structures may capture better the semantic relatedness among sentences.

Keywords: semantic similarity; paraphrase identification; heterogeneous network

1. Introduction

Paraphrase detection is a Natural-Language Processing (NLP) task that aims to automatically identify whether two sentences convey the same meaning. Bhagat and Hovy [1] define paraphrase as sentences or phrases that convey the same meaning using different wording. Moreover, these sentences represent alternative surface forms in the same language, expressing the same semantic content of the original forms [2].

Formally, a paraphrase may be modeled as a mutual (or bidirectional) entailment between a text T and a hypothesis H in the form, $T \rightarrow H$ and $H \rightarrow T$, that means T entails H and H entails T. For example, given a text T and a hypothesis H below, one may claim that they are paraphrases of each other, since $T \rightarrow H$ and $H \rightarrow T$.

T It is a strange term, but we have become used to it.
H This term is strange; however, we are accustomed to it.

According to Anchiêta and Pardo [3], studies that have focused on the paraphrase detection task for the Portuguese language (as defined above) are rare. A reason for this is the lack of large available corpora of paraphrases. As a consequence, there are few developed methods for paraphrase identification that could be useful for varied NLP tasks that might benefit from such knowledge, as semantic parsing [4], machine translation [5], automatic summarization [6], question answering [7] and plagiarism detection [8], among others.

To try to overcome such barriers, researchers developed and used the ASSIN corpus [9], which is focused on the textual entailment recognition task, but includes paraphrase examples. Formally, entailment recognition is the task of deciding whether the meaning of one text may be inferred from another [9].

The existing works that aim to detect paraphrase sentences in Portuguese [3,10], model this task as a machine-learning solution, building feature-value tables and training and testing classifiers. For a new sentence pair, features are computed and fed into the classifier to predict if the two sentences are paraphrases of each other. These approaches may suffer from two drawbacks. In the first one, the features may not capture well the semantics of the sentence pairs, producing unsatisfactory results. In the second, the authors apply sampling techniques to mitigate the unbalance issues of the ASSIN corpus, aiming to get more balanced data to improve the results of their models. These under- or over-sampling techniques may suffer from some shortcomings. In the over-sampling, the minority class can lead to model overfitting, introducing duplicate instances from a pool of instances that is already small [11]. On the other hand, in the under-sampling, the majority class can end up leaving out important instances that provide important differences between the two classes [12]. Other strategies that make use of synthetic data also suffer from criticism on the quality of the generated data.

To fulfill these gaps and explore other approaches for paraphrase detection, in this paper, inspired by Sousa et al. [13], we model the paraphrase detection task as a heterogeneous network. In this network, nodes represent tokens and sentence pairs, and the edges link the two node types. Networks/graphs have shown to be a powerful data structure that may capture well the relationship among the objects of interest [14].

Based on the network, we feed and train a classifier to predict if two sentences are paraphrases of each other. To evaluate our method, we use the ASSIN corpus. However, instead of applying a sampling technique to balance it, we adopt a back-translation strategy [15] for data augmentation to balance the data. This strategy maintains the original sentence pairs from the ASSIN corpus and add real sentences from another corpus with good translation quality. Our proposed method outperforms the best reported results, both in F-score and accuracy measures. Furthermore, the back-translation strategy helps to produce better models.

The remaining of this paper is organized as follows. Section 2 briefly presents the related work. In Section 3, we show the used corpora. Section 4 details our methodology to balance the ASSIN corpus and to model the problem. Section 5 presents the conducted experiments and obtained results. Finally, in Section 6, we conclude the paper, giving directions for future work.

2. Related Work

As pointed by Anchiêta and Pardo [3], few approaches strictly tackle the paraphrase detection task for the Portuguese language. Most of the research is on entailment identification that, according to Souza and Sanches [10], is different from paraphrase detection. Thus, following Souza and Sanches [10], we focus on the paraphrase detection task.

Consoli et al. [16] analyzed the capabilities of the coreference resolution tool CORP [17] for identification of paraphrases. The authors used CORP to identify noun phrases that may help to detect paraphrases between sentence pairs. They evaluated their method on 116 sentence pairs from the ASSIN corpus, achieving 0.53 F-score.

Rocha and Cardoso [18] modeled the task as a supervised machine-learning problem. However, they handled the issue as a multi-class task, classifying sentence pairs into entailment, none, or paraphrase. Thus, they employed a set of features of the lexical, syntactic, and semantic levels to represent the sentences in numerical values, and fed these features into some machine-learning algorithms. They evaluated their method on the training set of the ASSIN corpus, using both European and Portuguese partitions. The method obtained 0.52 of F-score using an SVM classifier.

Souza and Sanches [10] also dealt with the problem using a supervised machine-learning strategy. However, their objective was to explore sentence embeddings for this task. They used a pre-trained

FastText model [19] and the following features: the average of the vectors, the value of Smooth Inverse Frequency (SIF) [20], and weighted aggregation based on Inverse Document Frequency (IDF). With these features, their method reached 0.33 of F-score using an SVM classifier on balanced data of the ASSIN corpus for European and Portuguese partitions.

Cordeiro et al. [21] developed a metric named SUMO-METRIC for semantic relatedness between two sentences based on the overlapping of lexical units. Although the authors evaluated their metric on a corpus for the English language, the metric is language-independent.

Anchiêta and Pardo [3] explored the potentialities of four semantic features to identify paraphrase sentences. They computed the similarity of two encoded sentences as a graph using a semantic parser [22] and a semantic metric [23], the value of Smooth Inverse Frequency (SIF) [20], the cosine distance between two embedded sentences, and the value of the Word Mover's Distance (WMD) [24] between two embedded sentences. From these features, they trained an SVM classifier and obtained 0.80 F-score on the balanced ASSIN corpus.

For the English language, according to Mohamed and Oussalah [25], most of the research is categorized into three high levels, namely: corpus-based, knowledge-based, and hybrid methods. Here, in order to have a panoramic view of the achieved contributions for this language and to allow (indirect) comparisons to Portuguese state of the art, we briefly present the best results in the literature.

Mohamed and Oussalah [25] adopted a hybrid method addressing the problem of evaluating sentence-to-sentence semantic similarity when the sentences contain a set of named entities. The authors aimed to distinguish the computation of the semantic similarity of named entity tokens from the rest of the sentence text based on the integration of word semantic similarity derived from WordNet taxonomic relations [26], and named entity semantic relatedness inferred from Wikipedia entity co-occurrences and underpinned by Normalized Google Distance [27]. This approach reached 85.2% F-score on the MSRP corpus [28].

Ji and Eisenstein [29] adopted a corpus-based approach and used a distributional similarity model by designing a discriminative term-weight metric called TF-KLD. This metric outperforms the TF-IDF weighting scheme by re-weighting the sentence-term matrix in a different way. The authors evaluated their method on the MSRP corpus and achieved an 85.96% F-score.

Issa et al. [30] also adopted a corpus-based approach. For that, they combined Latent Semantic Analysis (LSA) [31] with Abstract Meaning Representation (AMR) [32] parsing. In this combination, the authors re-weighted the LSA sentence-term matrix according to a probability distribution over the AMR nodes, which was accomplish by means of the PageRank algorithm [33]. With this strategy, the authors reached 90% F-score on the MSRP corpus.

As we can see, our graph-based method is different from the literature both in the Portuguese and English languages. Moreover, our approach does not make use of external knowledge for computing sophisticated features: it uses only the sentences and tokens of the corpus.

3. Corpora

3.1. The ASSIN Corpus

ASSIN (Semantic Similarity Assessment and textual INference) [9] is a corpus with semantic similarity score and entailment annotations for sentence pairs in Portuguese, which was released for a shared task of the PROPOR 2016 Conference (http://nilc.icmc.usp.br/assin/). The corpus has 10,000 sentence pairs, half of which is written in Brazilian Portuguese and half in European Portuguese. Moreover, each language has 2500, 500, and 2000 pairs for training, development, and testing, respectively, as presented in Table 1.

Besides that organization, the sentence pairs of the corpus are labeled with inference classification, as entailment, none, and paraphrase, as shown in Table 2. One can see that the corpus is unbalanced concerning the paraphrase label, since the proportion of entailment and none labels is much higher than the paraphrase label.

Table 1. The ASSIN corpus.

Portuguese Language	Training	Development	Testing
Brazilian	2500	500	2000
European	2500	500	2000

Table 2. Frequency of the labels in the ASSIN corpus.

Label	Brazilian Port.			European Port.			Total #	Proportion %
	Train.	Dev.	Test.	Train.	Dev.	Test.		
Entailment	437	92	341	613	116	481	2080	20.80
None	1947	384	1553	1708	338	1386	7316	73.16
Paraphrase	116	24	106	179	46	133	604	6.04

3.2. The MSRP Corpus

MSRP (Microsoft Research Paraphrase) [28] is a widely used corpus for the paraphrase detection task for English. The sentence pairs are annotated in a binary strategy, i.e., as paraphrases or non-paraphrases. The corpus has 5801 sentence pairs, where 3900 are paraphrases and 1901 are non-paraphrases, as exposed in Table 3. As we can see, the MSRP corpus is unbalanced concerning the non-paraphrase label.

Table 3. Splits of the MSRP Corpus.

Label	Training	Testing	Total #	Proportion %
Paraphrase	2753	1147	3900	67.54
Non-paraphrase	1323	578	1901	32.46

In what follows, we detailed the developed methods for paraphrase identification and our strategy to mitigate the unbalance of the ASSIN corpus.

4. A New Method for Paraphrase Identification

4.1. Balancing the ASSIN Corpus

To formulate the paraphrase identification task as a binary classification problem, we did a modification to the ASSIN corpus, since it has three labels. Thus, we joined the entailment and none labels of the corpus into one unique label named "non-paraphrase", which is our negative class, as presented in Table 4.

Table 4. New configuration of the ASSIN corpus.

Label	Training	Development	Testing	Total #	Proportion %
Non-paraphrase	4705	930	3761	9396	93.96
Paraphrase	295	70	239	604	6.04

As we can see, the new configuration of the corpus is unbalanced concerning non-paraphrase label. Aiming to balance the corpus, we used the MSRP corpus. For that, we adopted a back-translation strategy [15] to translate the sentences of the MSRP corpus from English to Portuguese, as illustrated in Figure 1.

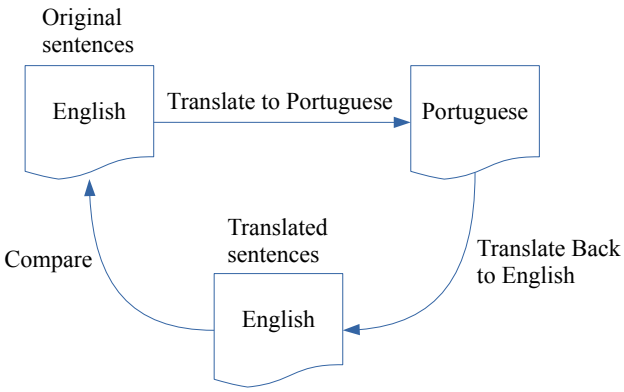

Figure 1. Back-Translation approach.

According to this figure, we translated the original sentences from the MSRP corpus to Portuguese, using the machine translation model provided by the Google Translate API (https://cloud.google.com/translate/). We adopted this model due to the good results achieved by other researchers [34,35] when using it. Next, we translate the Portuguese sentences back to English, because there are no Portuguese reference sentences to evaluate the quality of the translations. This way, we may measure the quality of the translations, comparing the original sentences (reference sentences) with the back-translated sentences (hypothesis sentences). To compute the quality of the translations, we calculated the harmonic mean between ROUGE (We used the F-score of the *ROUGE-L*.) [36] and BLEU [37] metrics, as in Equation (1).

$$F1 = 2 \times \frac{rouge \times bleu}{rouge + bleu} \tag{1}$$

We achieved 0.844 mean value (with a 0.08 standard deviation) when comparing the reference sentences to the hypothesis sentences. Taking into account the state-of-the-art results in machine translation [38], which achieves 35 points on the BLEU score, we may consider that the translated sentences from English to Portuguese have a good quality due to reached results both in the mean and the standard deviation. Thus, we used the translated sentences of the MSRP corpus to make the ASSIN corpus less unbalanced. Table 5 presents the ASSIN corpus plus the translated sentences of the MSRP corpus.

To create this less unbalanced corpus, we got the 2753 paraphrastic sentences from the training set of the MSRP corpus and put 2000 into the training set and 753 into the development set of the ASSIN corpus, respectively. As a consequence of these operations, the training set will have 2295 sentences and the development set will have 823 sentences. Moreover, we also got the 1147 paraphrastic sentences from the testing set of the MSRP corpus and put them into the testing set of the ASSIN corpus. As a result of that operation, the testing set will have 1386 sentences. After these procedures, we obtained a corpus in Portuguese with similar proportions in relation to the MSRP corpus.

Table 5. ASSIN corpus plus the translated sentences of the MSRP corpus.

Label	Training	Development	Testing	Total #	Proportion %
Non-paraphrase	4705	930	3761	9396	67.60
Paraphrase	2295	823	1386	4504	32.40

4.2. Modeling the Paraphrase Identification Task

To extract features, we first model the paraphrase identification task over a heterogeneous graph network. This network contains abundant information with structural relations (edges) among multi-typed nodes as well as unstructured contend associated with each node [39,40]. It has been used to automate feature engineering tasks, facilitating machine-learning tasks, and proving good

results for other tasks as helpfulness prediction, text classification and scientific impact measurement, among others [13,41–43].

We borrowed the formulation of a heterogeneous network from Chang et al. [42] and adapted it to our purpose. Our network may be viewed as a graph $G = (V, E)$, where $V = \{v_1, ..., v_n\}$ is a set of vertices and E is a set of edges, where an edge $e_{i,j} \in \{1, ..., n\}$ belongs to the set E if and only if an undirected and unweighted link exists between nodes i and j. Moreover, the graph G is associated with an object type mapping function $f_v : V \rightarrow O$, where O represents object sets and each node $v_i \in V$ belongs to one particular object type as $f_v(v_i) \in O$.

In our network, we defined two node types and two constraints. In the former, the nodes are tokens and sentence pairs, while in the latter the constraints are: (i) there is no link among token nodes and (ii) there is no link among sentence pair nodes. Thus, we link only token nodes with sentence pair nodes. We present the general scheme and an overview of the network in Figures 2 and 3, respectively.

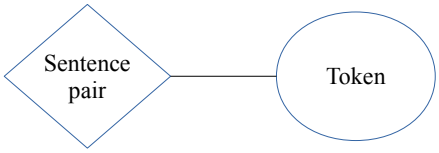

Figure 2. Scheme of the heterogeneous graph network.

As one can see, the edges are undirected and unweighted, and a sentence pair node may share several token nodes whenever the token is in the sentence pair, i.e., the edges between token nodes and sentence pair nodes are based on word occurrence in sentence pairs.

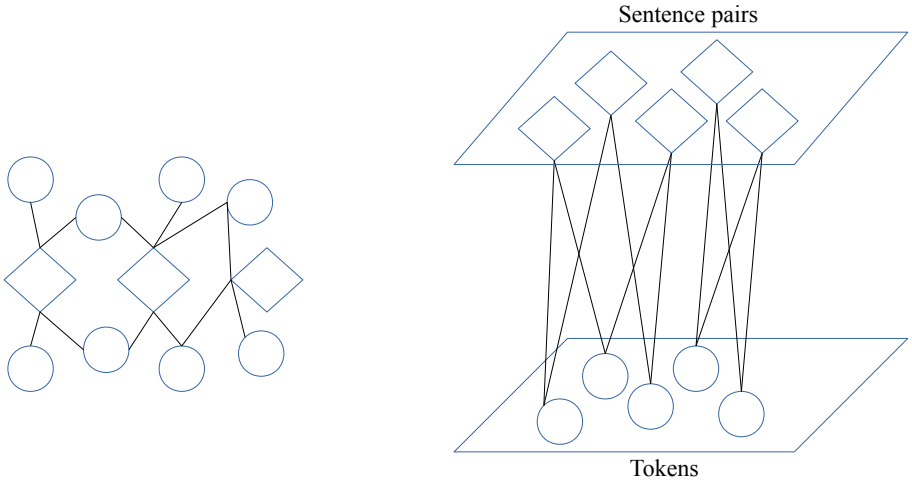

Figure 3. Overview of the heterogeneous graph network.

To extract the features regarding the network object classes, we applied a regularization method. Regularization is a kind of transductive classification method that aims to find a set of labels, minimizing a cost function and satisfying two conditions: (i) the method needs to be consistent with the set of labels manually annotated and (ii) the method needs to be consistent with the network topology, considering that nearest neighbors tend to have the same labels [14].

We tested three regularization methods: Gaussian Fields and Harmonic Function (GFHF) [44], Learning with Local and Global Consistence (LLGC) [45], and GnetMine [14]. The regularization algorithms require some nodes to be pre-labeled with their specific classes. Thus, one of the differences between these methods is if they modify these pre-labeled nodes. For example, the GFHF method does not modify the pre-labeled nodes, whereas the LLGC and GnetMine methods do. Besides, the GnetMine method works only on the heterogeneous networks, while the GFHF and LLCG methods work both on heterogeneous and homogeneous networks.

As a result, the regularization methods produce values related to coordinates for each object in the network, and these values may be used to feed a supervised machine-learning algorithm to learn and predict labels [46]. Table 6 presents an example of the output of a regularization method, where id is the identifier of the object, values refer to coordinates of each object in the network, and label 1 is a paraphrase, while label 0 is a non-paraphrase.

Table 6. Result of a regularization method.

Id	Value 1	Value 2	Label
0	0.002959	0.005986	1
29	0.002750	0.005323	1
54	0.003350	0.006426	0
69	0.003953	0.006793	0

4.3. Formulating the Paraphrase Identification Task

As before mentioned, we formulated the paraphrase identification task as a binary classification problem. For that, to train machine-learning algorithms, our method receives as input data the extracted features from a regularizer method in the form of $(x_1^{(i)}, x_2^{(i)}, b^{(i)})$ for $i \in [n]$, where n is the number of training sentences, $x_1^{(i)}$ and $x_2^{(i)}$ are the input sentences, and $b^{(i)} \in \{0, 1\}$ is the label, referring to a binary classification that informs whether $x_1^{(i)}$ and $x_2^{(i)}$ are paraphrases of each other. In summary, the aim is to learn a classifier c that, given unseen sentence pairs, i.e., a set of sentences S, classifies whether they are paraphrases, as in Equation (2).

$$c : S \times S \rightarrow 0, 1 \tag{2}$$

We tested four machine-learning algorithms, namely: Support Vector Machine (SVM), Naïve Bayes (NB), Decision Tree (DT), and Neural Network (NN). In what follows, we detail our experiments and the obtained results.

5. Experiments and Results

To evaluate our approach, we used the balanced ASSIN corpus with the translated sentences of the MSRP corpus, as depicted in Table 5. Moreover, as we commented before, we tested some classifiers from the Scikit-Learn library [47], as Support Vector Machine (SVM), Naïve Bayes (NB), Decision Trees (DT) and Neural Networks (NN), and we evaluated three regularization methods. Recall that the regularization methods require some nodes to be pre-labeled, so we ranged from 5% to 50% the number of pre-labeled nodes. The regularizers randomly pre-labeled the nodes. Supposing that the percentage of pre-labeled nodes is equal to 5%, it means that 0.25% of each class is randomly pre-labeled.

We achieved the best result with the LLGC regularizer, the NN classifier (We used a Multi-Layer Perceptron (MLP) with 2 hidden layers and 20 neurons in each hidden layer.), and 30% of the pre-labeled nodes on the balanced ASSIN corpus, as depicted in Table 7. It is important to highlight that only the training set is pre-labeled. The regularizer does not have access to labels of the testing set.

As we can see, from the 30% of pre-labeled nodes, both F-score and accuracy remain constants. We believe that the LLGC regularizer achieved the best results due to two properties. In the first place, it allows the pre-labeled nodes to be altered. This helps to correct errors in the labeling of nodes, improving the classification. In the second place, the algorithm decreases the excessive influence of objects with a high degree in the definition of the information of the nearest classes. This allows that the nodes get a different label from their neighbors.

Table 7. Results of the LLGC regularizer and the NN classifier.

Pre-Labeled (%)	NN		
	F-Score		Accuracy
	Paraphrase	Non-Paraphrase	-
5	0.00	0.84	0.73
10	0.00	0.84	0.73
15	0.70	0.90	0.85
20	0.71	0.90	0.85
30	**0.74**	**0.91**	**0.87**
40	0.74	0.91	0.87
50	0.74	0.91	0.87

We compared our best result with the works of Anchiêta and Pardo [3] and Souza and Sanches [10], since they also deal with the paraphrase detection task for Portuguese. Furthermore, we also compared our method with another graph-based method [48]. This method is a Graph Convolutional Network, which contains word nodes and document nodes. The number of nodes in the graph is the number of documents plus the number of unique words in a corpus. The edges among the nodes are based on word occurrence in documents and word co-occurrences in the whole corpus. Moreover, the weight of the edge between a document node and a word node is the Term Frequency-Inverse Document Frequency (TF-IDF) and the weight between two word nodes is the Pointwise Mutual Information (PMI) [49] value. Equation (3) summarizes the approach to weight an edge between nodes.

$$W_{i,j} = \begin{cases} \text{PMI}(i,j) & i,j \text{ are words, PMI}(i,j) > 0 \\ \text{TF-IDF}_{i,j} & i \text{ is document, } j \text{ is word} \\ 1 & i = j \\ 0 & \text{otherwiese} \end{cases} \tag{3}$$

In Table 8, we present the results of the comparison between the models, and, as we can see, our strategy outperformed the other methods, achieving better results than other models with only 30% of the pre-labeled data. It is important to say that we trained and evaluated these models on the balanced ASSIN corpus.

Table 8. Comparison among approaches on the balanced ASSIN corpus.

Method	F-score		Accuracy
	Paraphrase	Non-Paraphrase	-
Souza and Sanches [10]	0.45	0.55	0.50
Anchiêta and Pardo [3]	0.67	0.89	0.83
Yao et al. [48]	0.68	0.88	0.83
Ours	**0.74**	**0.91**	**0.87**

We further assessed whether the trained models on the balanced ASSIN corpus improve the results of the models when evaluated on the ASSIN corpus without balancing, i.e., we are interested in check if the data-augmentation strategy used to balance the ASSIN corpus contributes to improve the results when tested on the original ASSIN corpus. The results of this investigation are shown in Table 9.

One can see that all the models improved their results when trained on the ASSIN corpus with data augmentation, showing that the back-translation strategy for the paraphrase detection task is feasible to produce better models. For this experiment, the method of Anchiêta and Pardo [3] reached the best results, on average. Also, the graph-based methods performed poorly, having difficulty to

correctly predict a label with very few instances. ASSIN corpus has 239 instances as paraphrases and 3761 instances as non-paraphrases in the test set, as depicted in Table 4. To alleviate the difficulty of graph-based models to predict a label with very few instances, one may use boosting strategies, as RUSboost [50]. To investigate other approaches to tackle this subtlety remains for future work.

Table 9. Comparison among approaches on the ASSIN corpus.

Method	Data-Augmentation	F-Score	
		Paraphrase	Non-Paraphrase
Souza and Sanches [10]	✗	0.22	0.77
	✓	**0.28**	**0.80**
Anchiêta and Pardo [3]	✗	0.32	0.88
	✓	**0.35**	**0.92**
Yao et al. [48]	✗	0.00	0.89
	✓	**0.02**	**0.93**
Ours	✗	0.00	0.90
	✓	**0.07**	**0.98**

The balanced corpus and our graph-based model are available at (https://github.com/RogerFig/graph-paraphrase).

6. Final Remarks

In this paper, we presented a graph-based approach to model the paraphrase detection task. We defined a heterogeneous network with two semantic node types; sentence pairs and tokens. We created undirected and unweighted links among the token node and sentence pair node types. More than the method, we detailed a data-augmentation strategy, using back-translation to balance the dataset. Our approach outperformed the best results reported for paraphrase detection in Portuguese on the balanced ASSIN corpus both on accuracy and F-score measures.

The defined network is flexible and may be adapted to include other node types, as the embeddings of the sentences or tokens, for example. In addition, because of this flexibility, other network topologies may be explored, as creating weighted links between sentence pair nodes. Furthermore, heterogeneous networks may be applied to a broad range of other NLP tasks, as summarization, dependency parsing, sentiment classification, automatic essay scoring, and others. Another interesting future work is to verify which paraphrase types the systems detect. For that, one could follow the work of Kovatchev et al. [51] to annotate the paraphrase types that occur in the corpus. With this additional annotation layer, it may be possible to perform a qualitative evaluation, helping to explain which paraphrase types the models identify.

Author Contributions: R.T.A. helped in conceived the research, in the performed experiments, and writing the research. R.F.d.S. helped in development and writing of the research. T.A.S.P. supervised and helped writing the research. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by University of São Paulo Research Office grant number 668.

Acknowledgments: The authors are grateful to USP Research Office (PRP 668) and IFPI for supporting this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bhagat, R.; Hovy, E. Squibs: What Is a Paraphrase? *Comput. Linguist.* **2013**, *39*, 463–472. [CrossRef]
2. Madnani, N.; Dorr, B.J. Generating Phrasal and Sentential Paraphrases: A Survey of Data-Driven Methods. *Comput. Linguist.* **2010**, *36*, 341–387. [CrossRef]

3. Anchiêta, R.T.; Pardo, T.A.S. Exploring the Potentiality of Semantic Features for Paraphrase Detection. In Proceedings of the 14th International Conference on Computational Processing of the Portuguese Language, Evora, Portugal, 2–4 March 2020; Springer: Évora, Portugal, 2020; pp. 228–238.
4. Su, Y.; Yan, X. Cross-domain Semantic Parsing via Paraphrasing. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 9–11 September 2017; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 1235–1246.
5. Sekizawa, Y.; Kajiwara, T.; Komachi, M. Improving Japanese-to-English Neural Machine Translation by Paraphrasing the Target Language. In Proceedings of the 4th Workshop on Asian Translation, Taipei, Taiwan, 27 November–1 December 2017; pp. 64–69.
6. Jing, H.; McKeown, K.R. Cut and Paste Based Text Summarization. In Proceedings of the 1st North American Chapter of the Association for Computational Linguistics Conference, Seattle, WA, USA, 28 April 29–4 May 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2000; pp. 178–185.
7. Marsi, E.; Krahmer, E. Explorations in Sentence Fusion. In Proceedings of the Tenth European Workshop on Natural Language Generation, Aberdeen, UK, 8–10 August 2005; pp. 109–118.
8. McClendon, J.L.; Mack, N.A.; Hodges, L.F. The Use of Paraphrase Identification in the Retrieval of Appropriate Responses for Script Based Conversational Agents. In Proceedings of the 27th International Flairs Conference, Pensacola Beach, FL, USA, 21–23 May 2014; pp. 196–201.
9. Fonseca, E.R.; dos Santos, L.B.; Criscuolo, M.; Aluísio, S.M. Visão geral da avaliação de similaridade semântica e inferência textual. *Linguamática* **2016**, *8*, 3–13.
10. Souza, M.; Sanches, L.M.P. Detecção de Paráfrases na Língua Portuguesa usando Sentence Embeddings. *Linguamática* **2018**, *10*, 31–44. [CrossRef]
11. Weiss, G.M.; McCarthy, K.; Zabar, B. Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs? In Proceedings of the 2007 International Conference on Data Mining, Las Vegas, NV, USA, 25–28 June 2007; CSREA Press: Las Vegas, NV, USA, 2007; pp. 35–41.
12. Batista, G.E.; Prati, R.C.; Monard, M.C. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.* **2004**, *6*, 20–29. [CrossRef]
13. de Sousa, R.F.; Anchiêta, R.T.; Nunes, M.d.G.V. A Graph-Based Method for Predicting the Helpfulness of Product Opinions. *ISys-Rev. Bras. Sist. Inform.* **2020**, *13*, 1–16.
14. Ji, M.; Sun, Y.; Danilevsky, M.; Han, J.; Gao, J. Graph regularized transductive classification on heterogeneous information networks. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Barcelona, Spain, 20–24 September 2010; pp. 570–586.
15. Sennrich, R.; Haddow, B.; Birch, A. Improving Neural Machine Translation Models with Monolingual Data. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; Association for Computational Linguistics: Stroudsburg, PA, USA, 2016; Volume 1, pp. 86–96.
16. Consoli, B.S.; Neto, J.F.S.; de Abreu, S.C.; Vieira, R. Análise da capacidade de identificação de paráfrase em ferramentas de resolução de correferência. *Linguamática* **2018**, *10*, 45–51. [CrossRef]
17. Fonseca, E.; Sesti, V.; Antonitsch, A.; Vanin, A.; Vieira, R. CORP: Uma abordagem baseada em regras e conhecimento semântico para a resoluçao de correferências. *Linguamática* **2017**, *9*, 3–18. [CrossRef]
18. Rocha, G.; Lopes Cardoso, H. Recognizing Textual Entailment and Paraphrases in Portuguese. In *Progress in Artificial Intelligence*; Oliveira, E., Gama, J., Vale, Z., Lopes Cardoso, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2017; pp. 868–879.
19. Bojanowski, P.; Grave, E.; Joulin, A.; Mikolov, T. Enriching word vectors with subword information. *Trans. Assoc. Comput. Linguist.* **2017**, *5*, 135–146. [CrossRef]
20. Arora, S.; Liang, Y.; Ma, T. A simple but tough-to-beat baseline for sentence embeddings. In Proceedings of the 5th International Conference on Learning Representations, Toulon, France, 24–26 April 2017.
21. Cordeiro, J.; Dias, G.; Brazdil, P. A metric for paraphrase detection. In Proceedings of the International Multi-Conference on Computing in the Global Information Technology, Guadeloupe City, Guadeloupe, 4–9 March 2007; pp. 1–7.
22. Anchiêta, R.T.; Pardo, T.A.S. A Rule-Based AMR Parser for Portuguese. In Proceedings of the Advances in Artificial Intelligence—IBERAMIA 2018, Trujillo, Peru, 13–16 November 2018; pp. 341–353.

23. Anchiêta, R.T.; Cabezudo, M.A.S.; Pardo, T.A.S. SEMA: An Extended Semantic Evaluation Metric for AMR. (To appear) In Proceedings of the 20th International Conference on Computational Linguistics and Intelligent Text Processing, La Rochelle, France, 7–13 April 2019.
24. Kusner, M.; Sun, Y.; Kolkin, N.; Weinberger, K. From Word Embeddings To Document Distances. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; pp. 957–966.
25. Mohamed, M.; Oussalah, M. A hybrid approach for paraphrase identification based on knowledge-enriched semantic heuristics. *Lang. Resour. Eval.* **2019**, *54*, 457–485. [CrossRef]
26. Miller, G.A. WordNet: A lexical database for English. *Commun. ACM* **1995**, *38*, 39–41. [CrossRef]
27. Cilibrasi, R.L.; Vitanyi, P.M. The google similarity distance. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 370–383. [CrossRef]
28. Dolan, B.; Quirk, C.; Brockett, C. Unsupervised Construction of Large Paraphrase Corpora: Exploiting Massively Parallel News Sources. In Proceedings of the 20th International Conference on Computational Linguistics, Geneva, Switzerland, 23–27 August 2004; Association for Computational Linguistics: Stroudsburg, PA, USA, 2004; pp. 350–356.
29. Ji, Y.; Eisenstein, J. Discriminative Improvements to Distributional Sentence Similarity. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 891–896.
30. Issa, F.; Damonte, M.; Cohen, S.B.; Yan, X.; Chang, Y. Abstract Meaning Representation for Paraphrase Detection. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics, New Orleans, LO, USA, 1–6 June 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2018; pp. 442–452.
31. Landauer, T.K.; Foltz, P.W.; Laham, D. An introduction to latent semantic analysis. *Discourse Process.* **1998**, *25*, 259–284. [CrossRef]
32. Banarescu, L.; Bonial, C.; Cai, S.; Georgescu, M.; Griffitt, K.; Hermjakob, U.; Knight, K.; Koehn, P.; Palmer, M.; Schneider, N. Abstract Meaning Representation for Sembanking. In Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse, Sofia, Bulgaria, 8–9 August 2013; Association for Computational Linguistics: Stroudsburg, PA, USA, 2013; pp. 178–186.
33. Page, L.; Brin, S.; Motwani, R.; Winograd, T. *The PageRank Citation Ranking: Bringing Order to the Web*; Technical Report; Stanford InfoLab: Stanford, CA, USA, 1999.
34. Sobrevilla Cabezudo, M.A.; Mille, S.; Pardo, T. Back-Translation as Strategy to Tackle the Lack of Corpus in Natural Language Generation from Semantic Representations. In Proceedings of the 2nd Workshop on Multilingual Surface Realisation, Hong Kong, China, 3 November 2019; Association for Computational Linguistics: Stanford, CA, USA, 2019; pp. 94–103.
35. Cabezudo, M.A.S.; Inácio, M.; Rodrigues, A.C.; Casanova, E.; de Sousa, R.F. Natural Language Inference for Portuguese Using BERT and Multilingual Information. In Proceedings of the 14th International Conference on Computational Processing of the Portuguese Language, Evora, Portugal, 2–4 March 2020; Springer: Évora, Portugal, 2020; pp. 346–356.
36. Lin, C.Y. *ROUGE: A Package for Automatic Evaluation of Summaries*; Text Summarization Branches Out; Association for Computational Linguistics: Barcelona, Spain, 2004; pp. 74–81.
37. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; Association for Computational Linguistics: Philadelphia, PA, USA, 2002; pp. 311–318.
38. Edunov, S.; Ott, M.; Auli, M.; Grangier, D. Understanding Back-Translation at Scale. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–4 November 2018; pp. 489–500.
39. Sun, Y.; Han, J.; Yan, X.; Yu, P.S.; Wu, T. Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. *VLDB* **2011**, *4*, 992–1003. [CrossRef]
40. Zhang, C.; Song, D.; Huang, C.; Swami, A.; Chawla, N.V. Heterogeneous graph neural network. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Anchorage, AK, USA, 4–8 August 2019; ACM: New York, NY, USA, 2019; pp. 793–803.

41. King, B.; Jha, R.; Radev, D.R. Heterogeneous Networks and Their Applications: Scientometrics, Name Disambiguation, and Topic Modeling. *Trans. Assoc. Comput. Linguist.* **2014**, *2*, 1–14. [CrossRef]
42. Chang, S.; Han, W.; Tang, J.; Qi, G.J.; Aggarwal, C.C.; Huang, T.S. Heterogeneous network embedding via deep architectures. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, Australia, 10–13 August 2015; ACM: New York, NY, USA, 2015; pp. 119–128.
43. Dong, Y.; Chawla, N.V.; Swami, A. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD International Conference On Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; ACM: New York, NY, USA, 2017; pp. 135–144.
44. Zhu, X.; Ghahramani, Z.; Lafferty, J. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 912–919.
45. Zhou, D.; Bousquet, O.; Lal, T.N.; Weston, J.; Schölkopf, B. Learning with Local and Global Consistency. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004; pp. 321–328.
46. Bui, T.D.; Ravi, S.; Ramavajjala, V. Neural graph learning: Training neural networks using graphs. In Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Los Angeles, CA, USA, 5–9 February 2018; pp. 64–71.
47. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
48. Yao, L.; Mao, C.; Luo, Y. Graph Convolutional Networks for Text Classification. In Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 7370–7377.
49. Church, K.W.; Hanks, P. Word Association Norms, Mutual Information, and Lexicography. *Comput. Linguist.* **1990**, *16*, 22–29.
50. Seiffert, C.; Khoshgoftaar, T.M.; Van Hulse, J.; Napolitano, A. RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. *IEEE Trans. Syst. Man, Cybern.-Part A Syst. Hum.* **2010**, *40*, 185–197. [CrossRef]
51. Kovatchev, V.; Martí, M.A.; Salamó, M. ETPC - A Paraphrase Identification Corpus Annotated with Extended Paraphrase Typology and Negation. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation, Miyazaki, Japan, 7–12 May 2018.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).