

Lexical substitution using non-contextual and contextual word embeddings

Guillem Escriba (u188331), David Pérez (u188332) & Eduardo Ruiz (u162345)

Comparative table of results:

Model	Number of valid substitutes			Valid substitutes per context			% of contexts with valid substitutes			Overall Average Score**		
	K* = 5	K = 10	K = 50	K = 5	K = 10	K = 50	K = 5	K = 10	K = 50	K = 5	K = 10	K = 50
FastText	434	798	<u>2800</u>	0.57	1.047	<u>3.675</u>	35.6%	51.8%	<u>86.0%</u>	<u>7.1%</u>	6.8%	4.8%
Word2Vec	621	1159	<u>3076</u>	0.815	1.521	<u>4.037</u>	49.9%	66.7%	<u>84.1%</u>	<u>9.0%</u>	8.1%	4.1%
WordNet	1847			2.424			79.8%			8.1%		
XLM-RoBERTa	595	984	<u>2334</u>	0.781	1.291	<u>3.063</u>	47.2%	60.2%	<u>80.6%</u>	<u>8.5%</u>	6.8%	3.0%

Table 1:

*Is the number of similar words retrieved, for WordNet, all words are used.

** The Average Score of a substitute is the % of TRUE in the ground truth of this substitute and context. The Overall Average Score is the average of all the scores of all the contexts & substitutes.

Examples:

In this section it can be found examples of each model, due to the extension of each of those examples, we have chosen to keep here just one of each for readability (in the notebook there are 3 for each model). For benchmarking purposes the example is the same in all of the models to see how each model solves the problem and make comparisons. Those examples, even if they are not representative of the model, can be used to analyze the approach of the models. The performance of each model would be analyzed given the metrics instead of the examples.

1. FastText:

Target: press

Context: "Noriega has closed every independent newspaper, radio and television station and arrested, tortured or forced into exile a long list of reporters," the statement declared.

It added: "In Cuba, public enemy No. 1 of press freedoms in the hemisphere, repression of journalists both Cuban and foreign is worse than ever."

And in Nicaragua, promises of press freedom by the Sandinistas "have not materialized."

Substitutes of 'press':

journalist (70%), newspaper (60%), reporter (40%), presss (0%), press (0%), press (0%), release (0%), press (0%), medium (0%), thepress (0%), pressthe (0%), presser (0%), press (0%), release (0%), conference (0%), agentry (0%), junket (0%), pressthe (0%), pressroom (0%), button (0%), presser (0%), prss (0%), briefing (0%), apres (0%), conferece (0%), pressman (0%), relase (0%), publicity (0%), pree (0%), news (0%), conferance (0%), spacebar (0%), releasethe (0%), newsmedia (0%), buttion (0%), presss (0%), presse (0%), pressroom (0%), buton (0%), pressl (0%), flack (0%), prese (0%)

2. Word2Vec:

Target: press

Context: "Noriega has closed every independent newspaper, radio and television station and arrested, tortured or forced into exile a long list of reporters," the statement declared.

It added: "In Cuba, public enemy No. 1 of press freedoms in the hemisphere, repression of journalists both Cuban and foreign is worse than ever."

And in Nicaragua, promises of press freedom by the Sandinistas "have not materialized."

Substitutes of 'press':

journalist (70%), reporter (40%), medium (0%), hastily convene press (0%), breife (0%), news (0%), briefing (0%), conference (0%), presser (0%), secretary natalia timakova (0%), grape fermentation (0%), secretary jacquelyn hear (0%), shams ul safa aminzai (0%), press (0%), secretary greg massoni (0%), secretary jane jankowski (0%), spokesman yasuhide yamashiro (0%), secretary margita thompson (0%), secretary joanna doven (0%), pressconference (0%), chief barnabe gome (0%), secretary denise bottcher (0%), secretary alexei gromov (0%), advisor bishnu rijal (0%), press confrence (0%), kapihan sa senado (0%), confrence (0%), confrence (0%), secretary aaron mclear (0%), zhao yunshan director (0%), secretary marlin fitzwater (0%), postfight press (0%), confer ence (0%), pathrakarthara bhavan (0%), secretary tessa hafen (0%), secretary ben kieckhefer (0%), corby trouser (0%), jane jankowski daniels (0%), luca colajanni (0%), pressroom (0%), hadl auditorium (0%), secretary shawn feddeman (0%), agent tr reinman (0%), presss (0%), hurriedly convene press (0%), leiweke reunite (0%), kazakh mfa (0%), pressman (0%), beruaveau (0%), newser (0%)

3. WordNet:

Target: press

Context: "Noriega has closed every independent newspaper, radio and television station and arrested, tortured or forced into exile a long list of reporters," the statement declared.

It added: "In Cuba, public enemy No. 1 of press freedoms in the hemisphere, repression of journalists both Cuban and foreign is worse than ever."

And in Nicaragua, promises of press freedom by the Sandinistas "have not materialized."

Substitutes of 'press':

weigh (0%), squeeze (0%), compress (0%), exhort (0%), campaign (0%), agitate (0%), closet (0%), weight-lift (0%), pressing (0%), wardrobe (0%), urge_on (0%), iron_out (0%), imperativeness (0%), jam (0%), compact (0%), crusade (0%), printing_press (0%), iron (0%), public_press (0%), beseech (0%), military_press (0%), contract (0%), press (0%), insistence (0%), insistency (0%), push (0%), bid (0%), pressure (0%), entreat (0%), adjure (0%), fight (0%), urge (0%),

mechanical_press (0%), constrict (0%), press_out (0%), weightlift (0%),
conjure (0%), crush (0%)

4. XLM RoBERTa:

Target: press

Context: "Noriega has closed every independent newspaper, radio and television station and arrested, tortured or forced into exile a long list of reporters," the statement declared.

It added: "In Cuba, public enemy No. 1 of press freedoms in the hemisphere, repression of journalists both Cuban and foreign is worse than ever."

And in Nicaragua, promises of press freedom by the Sandinistas "have not materialized."

Substitutes of 'press':

journalist (70%), newspaper (60%), communication (30%), press (0%), medium (0%), all (0%), news (0%), democratic (0%), civil (0%), political (0%), the (0%), academic (0%), cultural (0%), our (0%), basic (0%), fundamental (0%), human (0%), religious (0%), most (0%), professional (0%), information (0%), free (0%), these (0%), editorial (0%), legal (0%), economic (0%), other (0%), journal (0%), many (0%), individual (0%), international (0%), public (0%), national (0%), press (0%), liberal (0%), scientific (0%), traditional (0%), personal (0%), digital (0%), social (0%), creative (0%), print (0%), independent (0%), critical (0%), speech (0%), various (0%), modern (0%), true (0%), common (0%), write (0%)

CONCLUSIONS

We have seen how well each model has performed using Table 1 and the notebook plots, this combined with the printed examples can give us a clear view of the behavior of each model with the benchmark.

On the one hand, by starting with a numerical analysis WordNet outperforms others in the number of valid substitutes and the percentage of contexts with valid substitutes, suggesting a strong grasp of vocabulary diversity and contextual applicability. FastText lags in these areas, indicating a narrower range of suggestions. Regarding FastText, while the basic model showed limitations in generating a diverse range of contextually appropriate substitutes, we experimented with an enhanced approach to better leverage its capabilities. by implementing the ‘get_enhanced_embedding’ method, we developed a technique that averages the embeddings of words within a contextual window around the target word. This modification to FastText, encapsulated in our ‘get_similar_words_context_fasttext’ function, aimed to bridge the gap between traditional non-contextual embeddings and the need for context. Interestingly, while Word2Vec shows superior overall average scores at lower K values, its performance declines at higher Ks, suggesting a trade-off between substitute quality and quantity. XLM-RoBERTa presents a balanced performance but experiences a notable drop in the context percentage and average score as K increases, hinting at challenges in maintaining contextual appropriateness with a broader set of substitutes. This synthesis of the numerical data suggests that while WordNet appears to offer a versatile range of substitutions, Word2Vec and XLM-RoBERTa might prioritize the relevance of fewer, more context-specific substitutions, particularly at lower levels of K.

On the other hand, by making use of the printed examples, we can see noticeable differences between the approaches of the models to filling the gap. For example, the WordNet, uses the word “press” as a verb instead of a noun leading to useless substitutes like “weigh”, “squeeze”, “compress”,... However as we have shown in a previous paragraph, the overall results of the WordNet are more than satisfactory, even more if we consider its drastically smaller execution time compared to the other models. FastText architecture that uses a smaller subwords lets it to correct lexical and grammatical mistakes as well as misspelled words, the problem is that there are not such errors in the benchmark so this capability only makes the model less precise as we can see in the substitute list with candidates such as “presss”, “prss” or “thepress”. The particular approach of the Word2Vec, letting predict multi-words such as “hastily convene press” or “secretary natalia timakova” is not very convenient in this case, even so, it maintains better results than FastText. Finally, the transformer model XLM RoBERTa is outperformed by the previous ones at least on metrics but looking closer to the examples, we can doubt how precise are those metrics since the transformer apparently provides more

meaningful words for these examples that might not be on the ground truth even if they are correct e.g. “democratic” or “civil”.

To sum up, the numerical and contextual analysis of the NLP models on the SWORDS benchmark reveals strengths and weaknesses inherent in each approach. WordNet shows commendable performance in terms of valid substitutes and context coverage, illustrating its strong vocabulary and contextual understanding, despite occasional mismatches in part-of-speech usage which can lead to irrelevant substitutes. FastText's proficiency in handling lexical nuances is unfortunately not leveraged within the benchmark's constraints, resulting in a lower precision reflected in the presence of improbable substitutes. Word2Vec, while showing promise in generating multi-word substitutes and maintaining high-quality results at lower K values, doesn't align well with the benchmark's requirement for single-word substitutes, indicating a misalignment with the benchmark's design. XLM-RoBERTa, though not leading in the metrics, offers meaningful substitutions that question the benchmark's sensitivity to the nuance and relevancy of substitutes that may not be in the ground truth set but are contextually appropriate. This suggests a possible limitation of the benchmark itself in evaluating the quality of substitutions beyond the predefined ground truth and highlights the importance of a multifaceted assessment that includes but is not limited to numerical metrics. Each model exhibits unique capabilities, underscoring the necessity of a tailored approach to model selection based on the specific demands of the task at hand.