Pontifical Catholic University
of Rio Grande do Sul

Measuring Semantic Similarity Between Sentences Using a Siamese Neural Network

Alexandre Y. Ichida - Felipe Meneguzzi - Duncan D. Ruiz

✉ alexandre.yukio@acad.pucrs.br, felipe.meneguzzi@pucrs.br, duncan.ruiz@pucrs.br

Motivation and Key Objectives

- Providing computers the ability to interpret the semantics of natural language is a nontrivial task, given its inherent ambiguity and multiple ways of expressing the same information.
- Encoding natural language texts in a semantic representation amenable to manipulation operations enable solving tasks such as eliminating redundant bits text or grouping text segments by their meaning
- The key objective of this work is to measure the semantic similarity between two sentences, capturing semantic features of natural language through a Siamese neural network architecture.

Siamese Neural Network Architecture

- Siamese neural networks are composed of two identical networks with independent inputs, using a distance function in the output layer in order to measure similarity a metric. [2].
- Our approach is based on the Siamese LSTM network defined by Mueller and Thyagarajan [3].
- We measure semantic similarity between two sentences using Siamese Neural Network composed of the following layers:
  - The input layer converts each word into a dense vector extracted from pre-trained *Word Embedding* models. These word vector representations encode semantic aspects from the words. [1].
  - We use GRU (*Gated Recurrent Units*) recurrent architecture in order to process word sequences, learning long-term dependencies.
  - The output layer uses a distance function which results in the difference between semantic representation of each sub-network.
- We trained our neural network using *Word Embeddings* of english words.

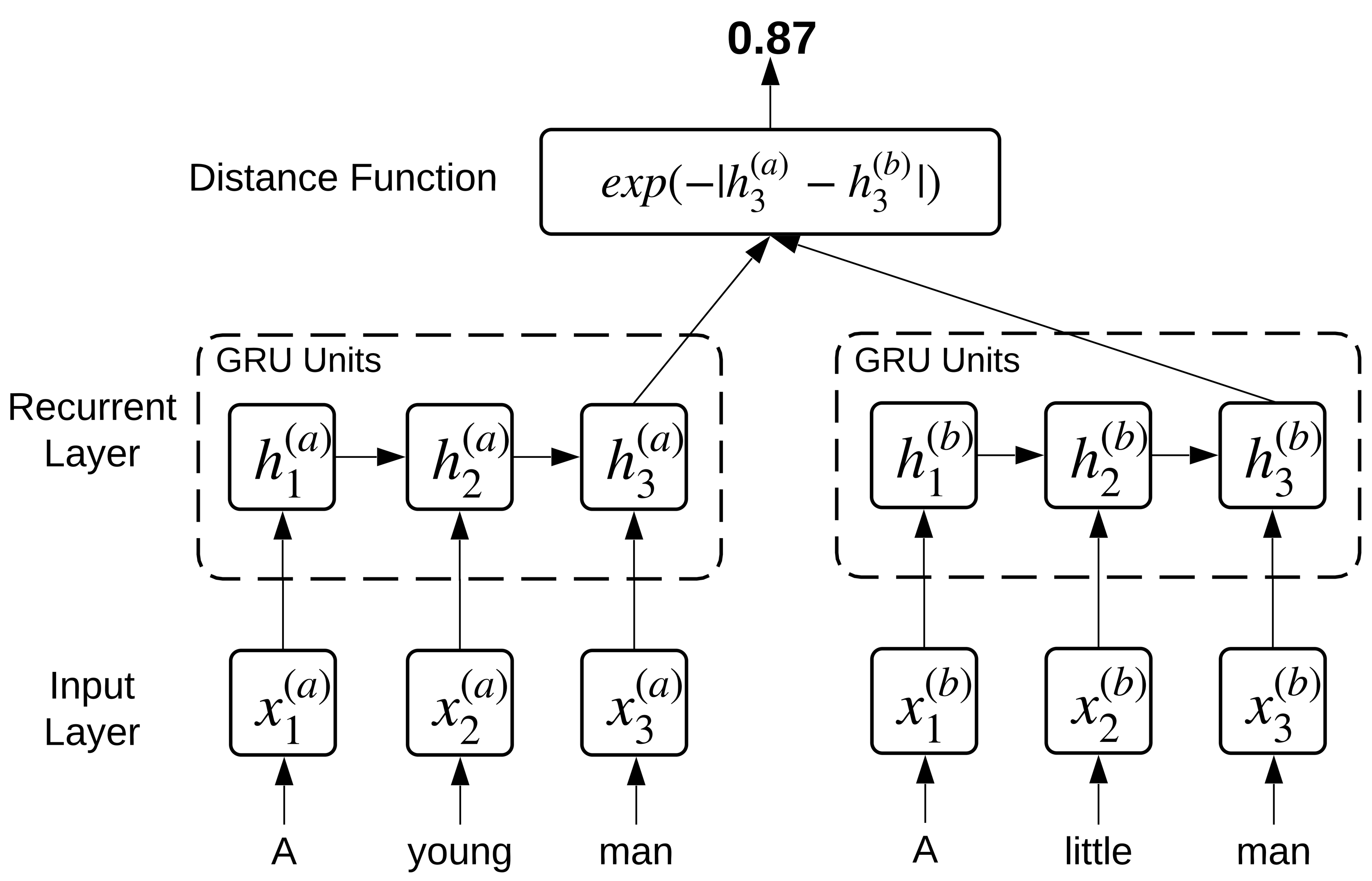

Figure 1: Diagram of our Siamese neural network architecture.

Dataset

- We use the *SICK (Sentences Involving Compositional Knowledge)* dataset which is available in SemEval *(International Workshop on Semantic Evaluation)*. SICK dataset contains 9927 sentence pairs with a 5000/4926 training/test split.
- Each sentence pair from the dataset is annotated with a numeric value [1,5] representing the degree of semantic relatedness between sentences.

Results

- Since the output of our model is a value between 0 and 1, we rescale the results to the same range as SICK dataset [1,5] in order to compare our model performance.
- The following table compares two specific sentences with a set of other sentences, with the semantic similarity value predicted from our model:

Sentence	Value
a woman is slicing potatoes	
a woman is cutting potatoes	4.79
potatoes are being sliced by a woman	4.41
tofu is being sliced by a woman	2.71
the woman was cut off from reality	1.88
two men are playing guitar	
the man is singing and playing the guitar	3.15
the man is opening the guitar for donations and plays with the case	2.91
two men are dancing and singing in front of a crowd	2.56

Related Work

- In order to compare our results with related work, we use metrics used in *SemEval* comparisons: Pearson correlation, Spearman correlation and Mean Square Error (MSE) between the predicted value and annotated value.
- A high correlation between sets of predicted values and annotated values means a superior performance in this comparison.
- The following table shows the comparison of related work which uses the same dataset.

Related Work	Pearson	Spearman	MSE
Illinois-LH (Lai & Hockenmaier - 2014)	0.7993	0.7538	0.3692
UNAL-NLP (Jimenez et al. - 2014)	0.8070	0.7489	0.3550
Meaning Factory (Bjerva et al. - 2014)	0.8268	0.7721	0.3224
ECNU (Zhao, Zhu, and Lan - 2014)	0.8279	0.7689	0.3250
Siamese-GRU (Ichida et al. - 2017)	0.8438	0.7844	0.3042
Dependency Tree-LSTM - StanfordNLP [4] (Tai, Socher, and Manning - 2015)	**0.8686**	**0.8047**	**0.2606**

Considerations

- Our model achieves results close to the state of the art using a neural network to learn a similarity metric, thus avoiding manual extraction of syntactic features of sentences.

Future Work

- Measure semantic similarity between sentences considering previous sentences in a corpus, thus improving context information of semantic representation.

References

[1] Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Jauvin. A neural probabilistic language model. *Journal of machine learning research*, 3(Feb):1137–1155, 2003.

[2] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. IEEE, 2005.

[3] Jonas Mueller and Aditya Thyagarajan. Siamese recurrent architectures for learning sentence similarity. In *AAAI*, pages 2786–2792, 2016.

[4] Kai Sheng Tai, Richard Socher, and Christopher D Manning. Improved semantic representations from tree-structured long short-term memory networks. *arXiv preprint arXiv:1503.00075*, 2015.