An Empirical Comparison Analysis on the Evolution of RNN Models using Multiple European Languages

Vikram Bhutani[a], Farshad Ghassemi Toosi[a]

[a] *Munster Technological University, Cork, Ireland*

Abstract

Humans are becoming globally connected more than ever. Communication is the key of this connection and languages are the means of this communication. We are witnessing globalization in the world that potentially cause some languages to dominate other languages and consequently the extinction of some languages. To be part of the globalized world and still to be able to communicate with own local language, language translation agents may be employed. Language Translation (LT) as an (Natural Language Processing) NLP application is a short answer to the aforementioned problem. In the last few decades we have seen the development of multiple Machine Learning techniques in language translation. In this work, we are presenting a new Recurrent Neural Network (RNN) architecture/model and experimenting a set of comparison analyses between our model and four existing RNN models (Simple RNN, Bidirectional RNN, Embedded RNN and Seq2Seq). The experiment is performed on an open-source repository for four pairs of languages: 1) *English* to *Irish/Gaelic*, 2) *English* to *Spanish*, 3) *Irish/Gaelic* to *English* and 4) *Spanish* to *English*. Our result indicates that on average our proposed model outperforms all the other models for all four pairs of languages. The result also indicates that models, on average, favour the pairs of languages where English is the source language. *English* to *Spanish*, on average, has the highest performance compared to other pairs and *Irish/Gaelic* to *English*, on average, has the lowest performance.

Keywords

RNN, Bidirectional, Seq2Seq, Irish Language, Language Translation.

CERC 2021: Collaborative European Research Conference, September 09–10, 2021, Cork, Ireland

vikram.bhutani@intel.com (V. Bhutani); Farshad.Toosi@mtu.ie (F.G. Toosi)
0000-0002-0877-7063 (V. Bhutani); 0000-0001-7116-9338 (F.G. Toosi)

1. Introduction

There are nearly 100 alive languages in the world. Each language has a set of different features (e.g., grammar, phonetic, structure and alphabet). The differences between these languages and their unique structures, make the task of automated translation, specific to the source and target languages. Therefore, developing and training a specific model with a known set of hyper parameters that can be generalized for several different languages if not impossible, is a complicated task. Even when for a fixed pair of languages (e.g., English and Arabic), it is still not easy to come up with a unique model and a fixed set of hyper parameters that can be used for every type of content (e.g., literature translation, technical translation, speech translation and etc) [1]. Deep learning techniques [2] as a subset of Machine Learning techniques are the most commonly used family of techniques in the area of machine translation. Most of the current translation technologies make use of Recurrent Neural Networks (RNN) for the task of language translation.

Recurrent Neural Network (RNN) is a sub-class of Artificial Neural Network that is comprised of a directed graph where the output is used again as a new input. Traditional neural networks (non-RNN models) have an input layer with a certain number of neurons that makes the network to be compatible with input with a certain size. This can be seen as a limitation when the input data has no pre-defined size. Examples of such cases is textual input where each input is a sentence with a number of words and sentences may have different size.

One of the objectives of the Recurrent Neural Network idea was to address the mentioned issue so that inputs with different sizes may be fed to the network [3]. The initial success of RNN made researchers to develop different variations of RNN model to deal with specific types of problems.

In this work, a new hybrid RNN-based model is presented. This new model takes advantage of existing RNN models and its performance is compared against other well-known RNN variations. Three European languages (English, Spanish and Irish/Gaelic) are employed for the comparison analysis task:

We mainly aim to address the following three questions:

- Although all these three languages are European languages, is there a noticeable performance difference among these pairs of languages under the same model?
- How model's performance is affected by reversing the source and target languages e.g., (English to Spanish) vs (Spanish to English).
- Each RNN model has a particular architecture with a set of layers, how to take advantage of different models' architecture in one single model (**Hybrid Model**, the main contribution of this work).

2. Literature Review

One of the limitations of basic neural networks is the requirement for the input to have a fixed or predetermined size. Recurrent Neural Networks (RNNs) advanced the basic neural networks by introducing a new structure where inputs do not have to have a fixed or predetermined size. Applications (e.g., language translation) where the input data does not or can not have a fixed size would greatly benefit from this type of architecture.

In the last few decades, there have been a number of RNN models developed for language translation such as *Sequence2Sequence* [4] and yet the research in this area is still active. For instance, [5], discussed the problem of multi language translation using encoder-decoder and allocating multiple decoders to multiple target languages. This way, they tried to have a simultaneous translation from language A (encoder) to languages B, C and so on (decoders). In another work, [6] proposed two novel models based on Recurrent Neural Networks for three different pairs of languages: German-English, Arabic-English and Chinese-English. Their novel approaches in Recurrent Neural Networks has two variations: a word-based model and the other one is phrase-based. The models indicate an improvement compared to the base models based on two metrics (BLEU and TER). Hu et al [7] also proposed a novel technique, MTU (Minimum Translation Unit) based approach against the classical n-gram back-off model on WMT 2012 French-English dataset. Their evaluation metric is BLEU and resulted in 0.8 improvement compared to traditional n-gram model.

Traditional RNN models like Seq2Seq are the model of choice for most of the NLP applications but training such models using big data can be very challenging. New technologies that use transformers with self and multi-head attention are proving to be state-of-the-art that can significantly reduce computational requirements [8]. The transformers use attention mechanism thus require less data computations and are less expensive as opposed to traditional RNN which uses LSTM (Long Short Term Memory) or GRU(Gated Recurrent Unit) [9].

There have been major developments in the field of NLP where RNN and attention algorithms are used together to achieve high accuracy and faster training times like in XLNet architectures [10]. Major difficulties when applying Transformer to language translation applications is that it requires more complex configurations(e.g., optimizer, network structure, data augmentation) than the conventional RNN based models. Recent studies shows that RNN models using global or local attention mechanism techniques can be used as the state-of-the-art solution [11].

Table 1
English-Irish-Spanish translation dataset

Details	English text	Irish text	Spanish text
Total Number of sentences	138460	138460	138460
Maximum sentence length	21	25	26
Total Number of words	1555241	2147273	1636051
Vocabulary size	679	1013	912

3. Dataset

In order to have a more comprehensive analysis, an open-source dataset is employed from *WMT14* repository published by *STATISTICAL MACHINE TRANSLATION* [12]. This dataset is currently available in many European languages such as *English*, *German*, *French* and *Czech*. The topic of the dataset is related to many widely used AI applications and the dataset has been mostly used for machine learning performance indicators and NLP benchmarks. Since the aim of this work was to work on Gaelic and Spanish languages, respective translations by services which include python APIs by Google [13] and Ai translate services [14] is performed.

The features of each language after the translation (English to Spanish using DeepL) are displayed in Table 1.

Like any other Machine Learning problems, the language translation requires pre-processing as well. Since languages have different types of symbols, features and structure, the task of pre-processing can be challenging and can affect the performance significantly. The textual data should be carefully examined, properly cleaned and transformed appropriately before feeding them into the RNN models. There are many stages involved in NLP data processing before it is fed into the translation models. All three data-sets are undergoing the following standard processes:

- Data cleaning: Below are some of the most frequent types of noise that is present in text data:
  1. Unicode and other symbols include removal of special characters such as "’
  2. Removal of html tags.
  3. Removing numbers, generally numbers are not required to be translated as they are generic.
  4. Links: Links can be of many forms and most of them consist of strange symbols or short-codes that can present in the document.
- Tokenization: splitting our text into minimal meaningful units for ML algorithm to understand the data in the numerical form.

Table 2
The Details of all variants of RNN Models.

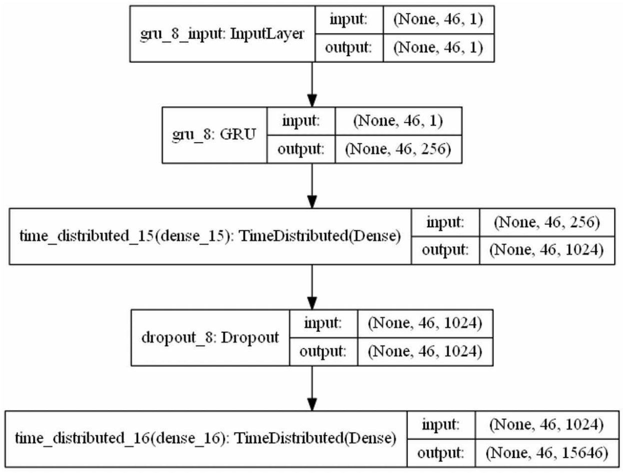
Model's Details	Simple RNN	Bidirectional RNN	Embedded RNN	Seq2Seq RNN	Hybrid RNN
Activation	relu	relu	relu	relu	relu
Final Layer Activation	softmax	softmax	softmax	softmax	softmax
Dropout	0.4	0.3	0.4	0.4	0.3
Layers	4	4	5	6	7
Optimiser	Adam	Adam	Adam	Adam	Adam
Learning rate	0.001	0.0001	0.001	0.001	0.001
Batch size	128	128	128	128	128
Recurrent Regularizer	0.000001	0.000001	0.000001	0.000001	0.000001
Epsilon	1e-08	1e-08	1e-08	1e-08	1e-08
Kernel Regularizer (decay rate)	1e-6	1e-6	1e-6	1e-6	1e-6

- Normalisation: Normalization is one of the important pre-processing steps and its attempt is to make a single representation of words with multiple representations. Stemming and lemmitising are key steps here.

- Stop words removal: Removing stop-words is another essential step. The main reason for stop-word removal is that these stop-words generally do not add new information to the text but are just a language construct.

- Embeddings and Representations. Once the dataset is cleaned, the text is converted into some kind of numerical representation to make them understandable for Machine Learning where they only understands numbers.

- Sentences padding: Proper padding is added to the sentences so that it will keep sentences to same size before the tensor multiplications is performed. This also helps in computational of high dimensional tensors. The <start> and <end> tokens are also added in each sentence to mark the start and end of sentences for tokenazation.



Simple RNN Model Layers

Figure 1: Simple Model.

4. Methodology

In this work, a comparative translation analysis is performed on different pairs of languages using five different models.

Four different pairs of languages using three European languages (English, Irish/Gaelic and Spanish) are employed in the experiment as follows:

- English to Gaelic language translation.
- Gaelic to English language translation.
- English to Spanish language translation.
- Spanish to English language translation.

As part of this research, we also aim to find the optimal hyper parameters for the best performing model and final selection will be done based on the performance metric.

4.1. Models

All the employed models in this work are from the family of RNN architecture and the details of the models' architectures are illustrated in Figure 1. The hyper parameters of every single model were decided experimentally. The main contribution of this work is the Hybrid RNN model that is illustrated in Figure 1, the architecture has components/layers from three other models: Embedded, Bidirectional and Seq2Seq.

Each model has a set of hyper parameters such as learning rates, optimizer, dropout and etc; the details of the hyper parameters are displayed in Table 2.

The hyper parameters are decided experimentally. For Learning Rate, seven values are tested (0.01, 0.001, 0.009, 0.0001, 0.00001, 0.000001 and 0.0000001) and the winner is 0.001, See Table 2 . A similar experimental approach is employed to decide an optimizer for the models. Six different optimizers (*adagrad*, *adam*, *SGD*, *adadelta*, *rmsprop* and *adamax*) are tested and based on

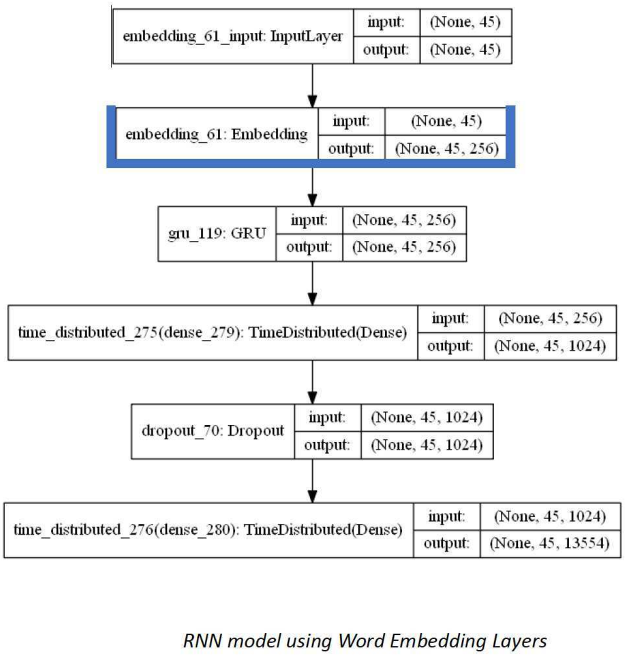


Figure 2: Embedded Model.

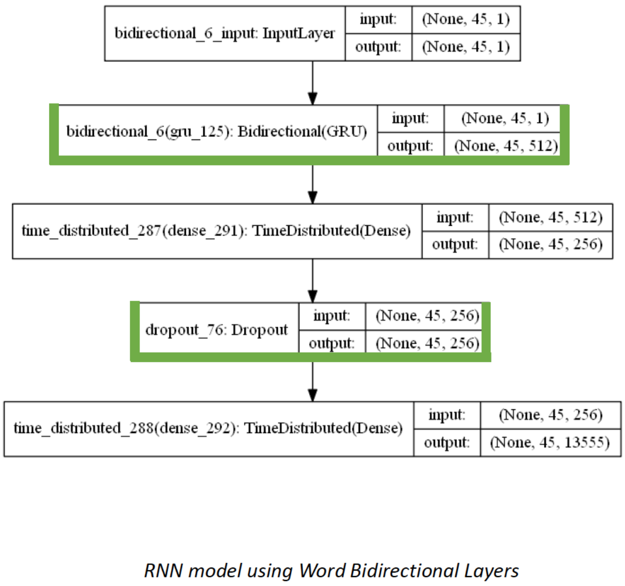
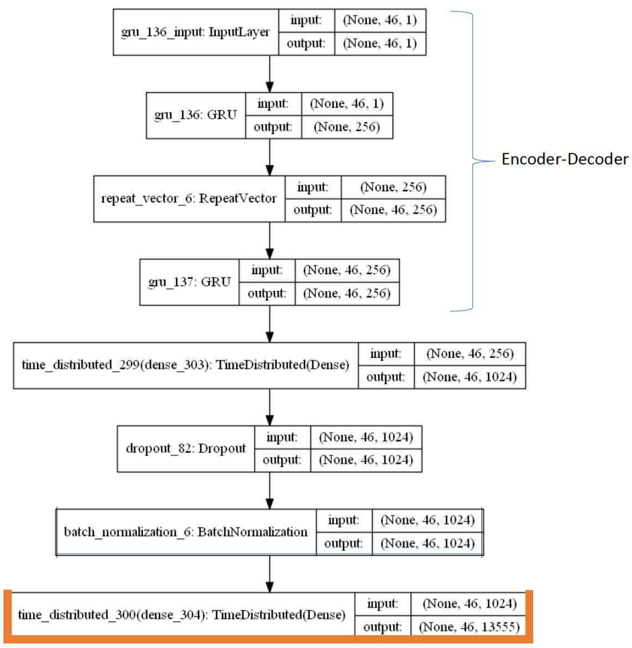


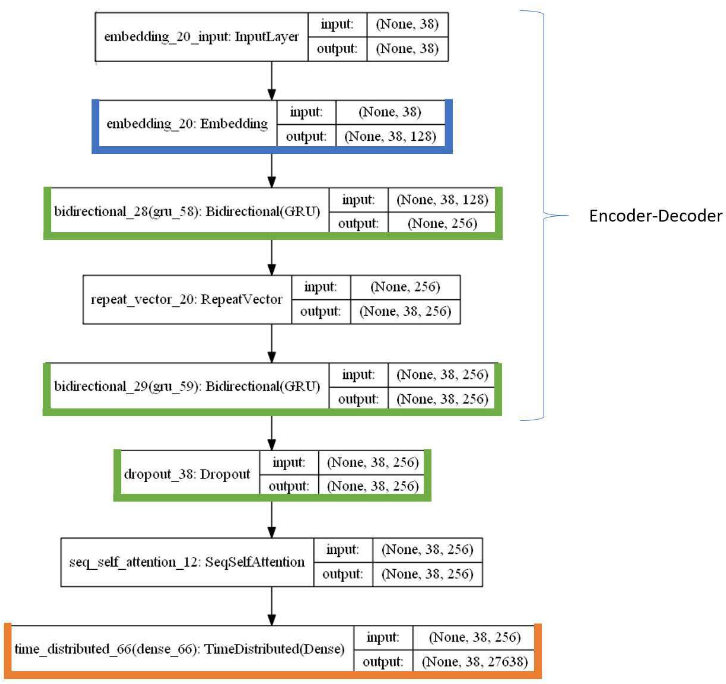
Figure 3: Bidirectional Model.

the accuracy performance, *adam* optimizer is selected. Although *adadelta* optimiser performed well in certain language pairs, *adam* optimizer is decided for all models due to its consistency across all language pairs.



Seq2Seq RNN model

Figure 4: Seq2Seq Model.



Final RNN model

Figure 5: Hybrid Model.

5. Results

The results of the whole experiment is divided into five different sub-experiments using five different models as follows: 1) Simple RNN [15] 2) Embedded RNN [16] 3) Bidirectional RNN [17] 4) Seq2Seq [18] and 5) Hybrid RNN model. All the examined models in this work are variations of RNN (Recurrent Neural Network) on two pairs of languages reciprocally: 1) *(Gaelic/Irish* $\Longleftrightarrow$ *English)* 2) *(Spanish* $\Longleftrightarrow$ *English)*. The following sections describe the details of each model and their results individually.

5.1. Simple RNN

The first experiment is carried out by a simple RNN [15] with 5 layers (details can be found in Table 2 and Figure 1. Simple RNN comprises of layer with GRU (Gated Recurrent Unit) followed by dense time distributed layer using softmax activation. This model is not deep enough and using minimal number of trainable parameters.

Figure 8 and 9 show the accuracy and loss distribution using simple RNN model on two pairs of languages reciprocally: *Gaelic to English*, *English to Gaelic*, *Spanish to English* and *English to Spanish*. The highest validation accuracy belongs to the *English to Spanish* pair followed by *Spanish to English* however the highest BLEU score belongs to *English to Gaelic* followed by *Gaelic to English*, See Table 3.

5.2. Embedding RNN

The first layer of the Embedding model [16] is word Embedding layer with the target vocabulary size. Word Embedding layer converts words into the dense vectors and helps in understanding the context of a word so that similar words have similar embeddings.

The highest validation accuracy for this model is resulted on *English to Spanish* pair followed by *English to Gaelic* and then *Spanish to English*. The highest BLEU score also belongs to *English to Spanish* and *Spanish to English* followed by (with relatively large gap) *English to Gaelic* and *Gaelic to English*

The performance of the validation and training sets are quite close to each other for *Spanish* $\Longleftrightarrow$ *English* pairs as opposed to *Gaelic* $\Longleftrightarrow$ *English* pairs, See Figures 12 and 13.

5.3. Bidirectional RNN

The Bidirectional model [17] includes Bidirectional Recurrent layers. The number of units in Bidirectional layer is doubled which resulted in more trainable parameters and thus increasing computational cost. While in simple RNN there is a single GRU layer, in Bidirectional model there are two LSTM or GRU cells activated to support forward and backward propagation. This can provide additional context to the network and result in faster learning of the model [17].

The highest validation accuracy by this model belongs to the *English to Spanish* followed by *Spanish to English* and *English to Gaelic*. *English to Spanish* is also the winner for BLEU score followed by *Spanish to English* and *English to Gaelic*.

Table 3
All Language Translations Performance Quality

Model	Simple RNN	Embedded	Bidirectional	Seq2Seq	Hybrid RNN Model
English to Gaelic					
Training Accuracy	80%	90%	85%	91%	95%
Validation Accuracy	78%	84%	82%	86%	89%
BLEU score	0.22	0.31	0.30	0.32	0.33
Epoch time (s)	140s	180s	220s	180s	190s
Gaelic to English					
Training Accuracy	77%	80%	81%	82%	83%
Validation Accuracy	76%	79%	79%	80%	82%
BLEU score	0.19	0.30	0.29	0.27	0.36
Epoch time (s)	120s	180s	260s	180s	210s
English to Spanish					
Training Accuracy	83%	88%	92%	95%	97%
Validation Accuracy	84%	87%	93%	93%	97%
BLEU score	0.16	0.31	0.33	0.27	0.42
Epoch time (s)	150s	220s	270s	200s	250s
Spanish to English					
Training Accuracy	81%	84%	82%	84%	85%
Validation Accuracy	81%	83%	82%	83%	84%
BLEU score	0.2	0.40	0.30	0.25	0.40
Epoch time (s)	150s	220s	290s	180s	260s

Spanish $\Longleftrightarrow$ *English* pairs seem to have similar performance on the validation and training sets compared to *Gaelic* $\Longleftrightarrow$ *English*, See Figures 10, 11 and Table 3.

5.4. Seq2Seq RNN

The Seq2Seq [18], also known as Encoder-Decoder, is the forth experimented model in this work. This model is the most popular model that is widely used in Autoencoders, Variational Autoencoders and in RNNs. We have also added self attention in this model to draw global dependencies between inputs and outputs [9].

While the highest validation accuracy under this model belongs to *English to Spanish* followed by *English to Gaelic*, the highest BLEU score belongs to *English to Gaelic* followed by *English to Spanish*.

Figures 14 and 15 show the accuracy and loss distributions using Seq2Seq model. A similar pattern to the previous models emerge here as well. The validation and training accuracy for *Spanish* $\Longleftrightarrow$ *English* pairs seem to be so close to each other while there is a considerable difference between the validation and training accuracy for *Gaelic* $\Longleftrightarrow$ *English* pairs.

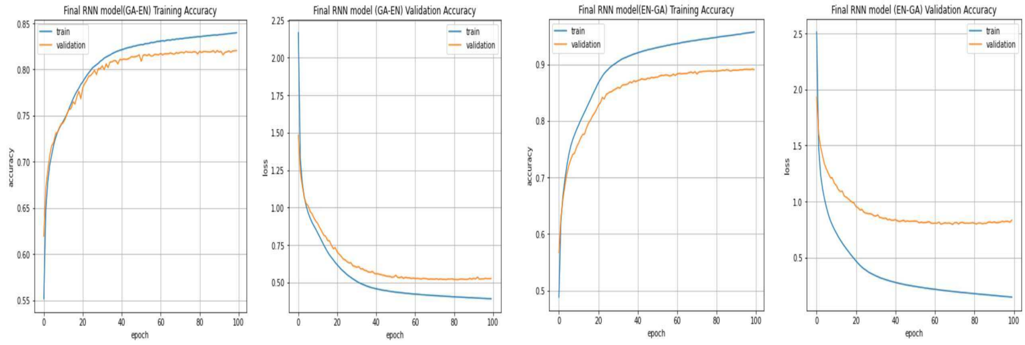


Figure 6: Hybrid RNN Model using Gaelic and English Pair

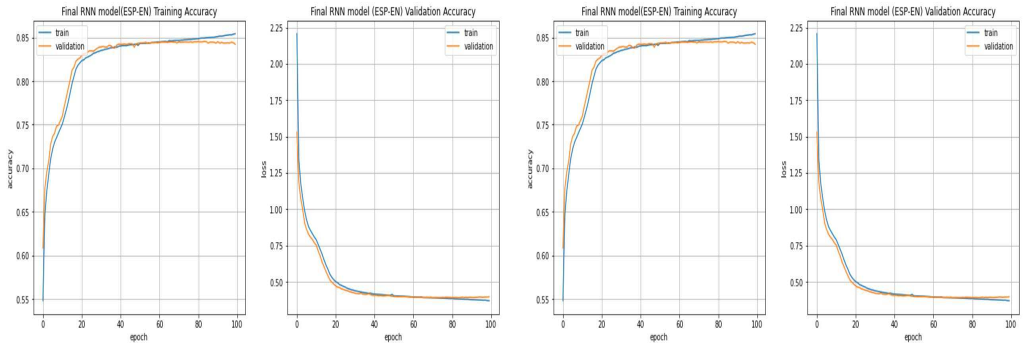


Figure 7: Hybrid RNN Model using Spanish and English Pair

5.5. Hybrid RNN Model

Hybrid RNN model is the main contribution of this work. Our proposed model contains features from three existing models (Embedding, Bidirectional and Seq2Seq). As shown in Table 3, Hybrid model is the winner in all the experiments and for all the metrics e.g., Validation and Training accuracy and BLEU score. Although Hybrid model has the highest performance compared to the previous models, its running time is computationally heavier as oppose to other models. As illustrated in Figures 6 and 7, the hybrid model converges faster for *Spanish* $\Longleftrightarrow$ *English* pairs as opposed to *Gaelic* $\Longleftrightarrow$ *English* pairs. The table 3 summarises the training and validation scores of all five employed neural networks for four different pairs of languages. The same table also shows the BLEU accuracy scores achieved by different models. The detailed comparison of variants of RNN models used in evaluation is shown in Figure 16. The result indicates that the performance of Hybrid model outperform all other RNN variants.

Figure 17 shows the different settings that are applied for model's hyperparameters and corresponding model performance. The model used in hyperparameter tuning is Hybrid RNN model. Based on the hypothesis and computational constraints, it is assumed that 100 epochs are enough to conclude the optimal hyperparameter values. The final distribution clearly shows the optimal learning rate, optimiser and regularisation that should be applied to the deep neural network. Based on the results, optimal settings of hyperparameter values are selected to maximize performance.

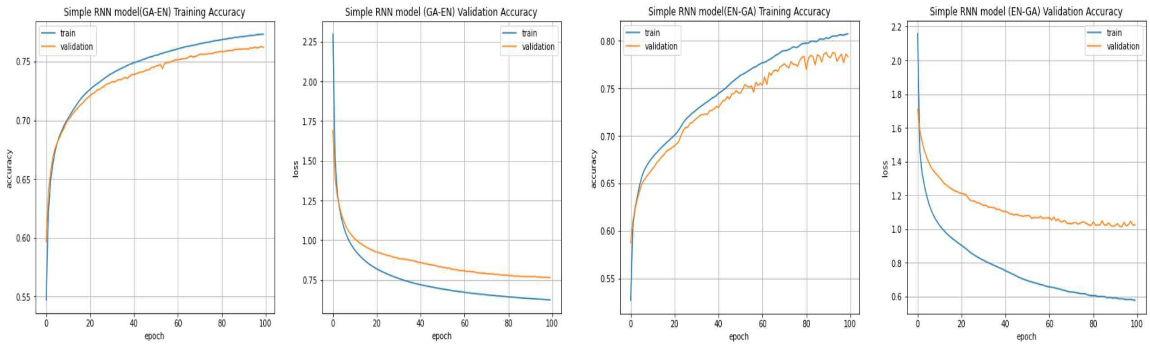


Figure 8: Simple RNN Model using Gaelic and English Pair

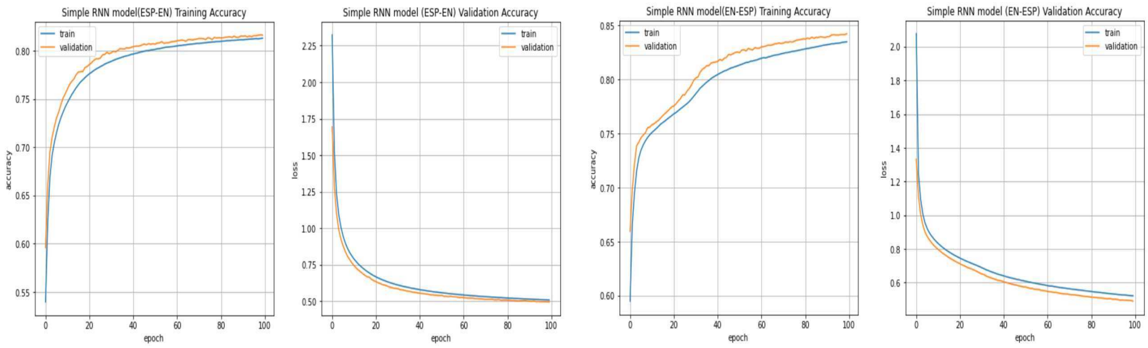


Figure 9: Simple RNN Model using Spanish and English Pair

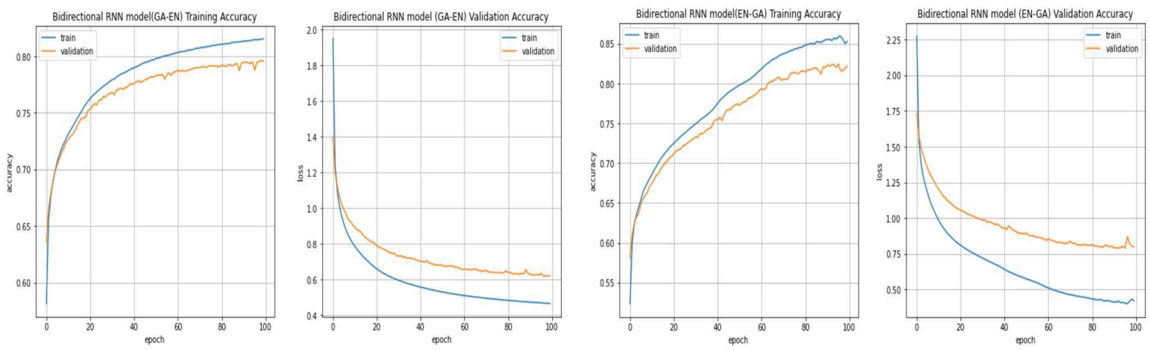


Figure 10: BiDirectional RNN Model using Gaelic and English Pair

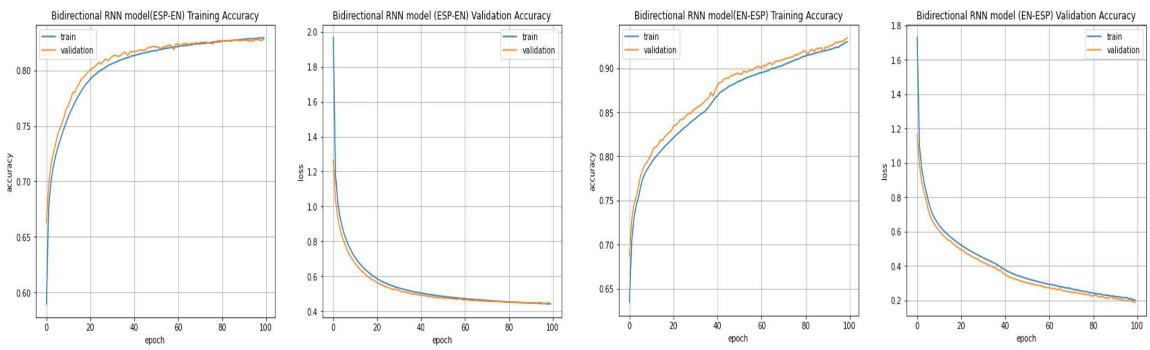


Figure 11: BiDirectional RNN Model using Spanish and English Pair

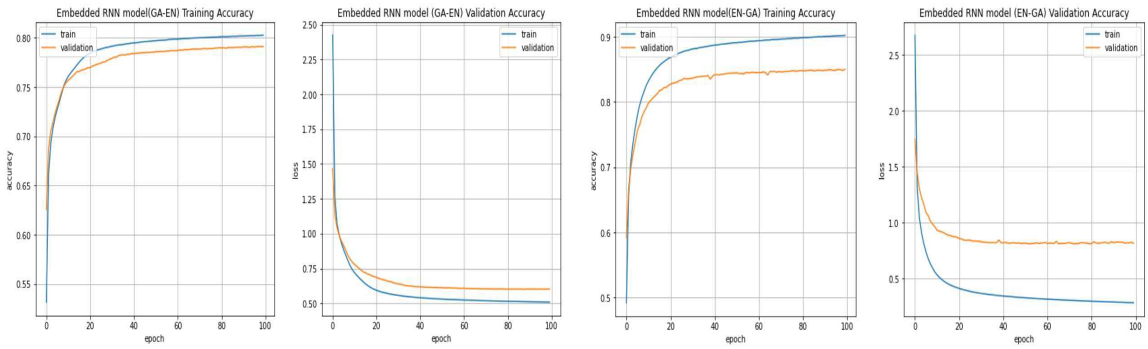


Figure 12: Embedded RNN Model using Gaelic and English Pair

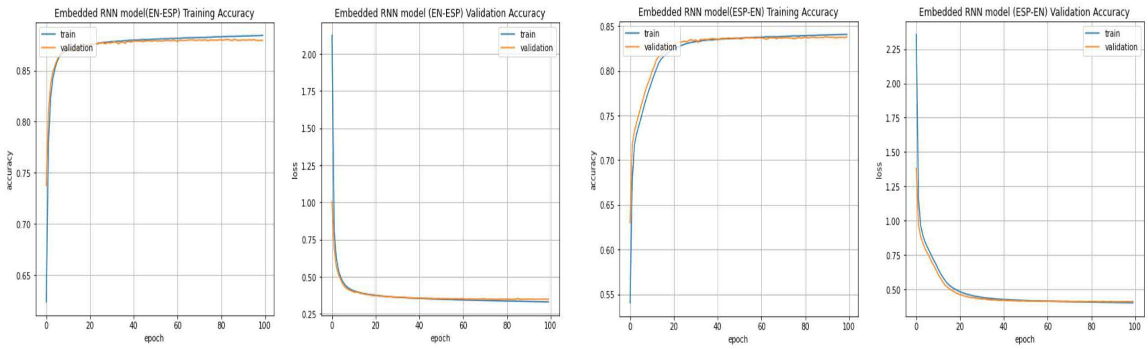


Figure 13: Embedded RNN Model using Spanish and English Pair

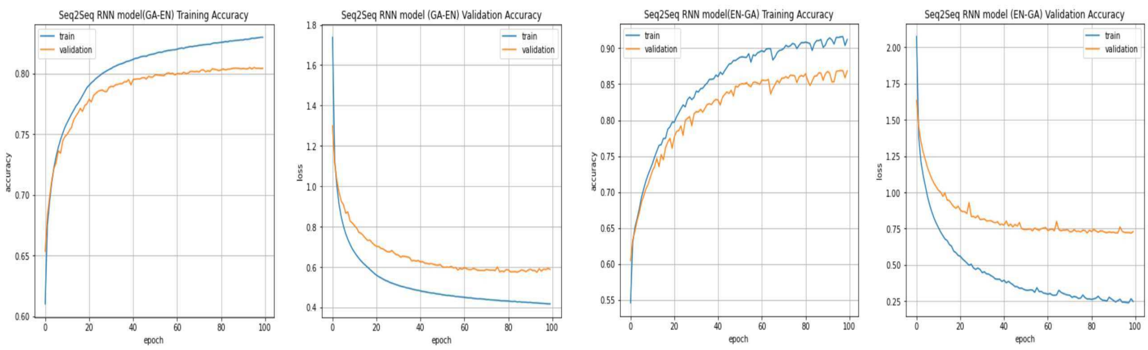


Figure 14: Seq2Seq RNN Model using Gaelic and English Pair

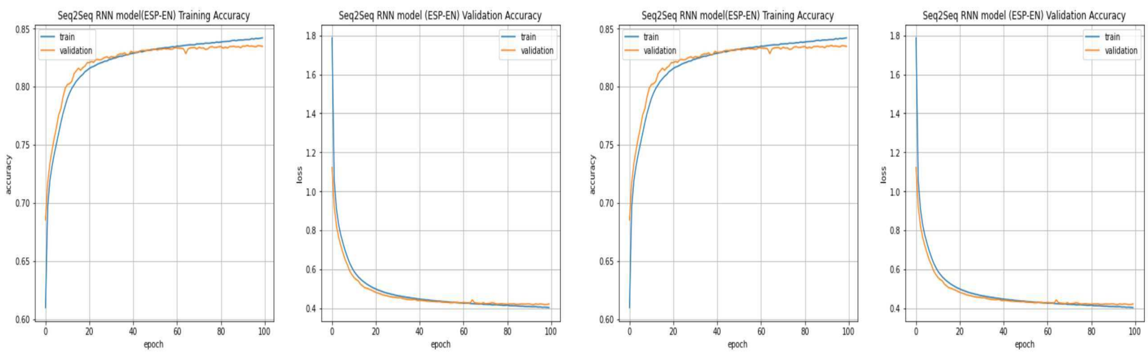


Figure 15: Seq2Seq RNN Model using Spanish and English Pair

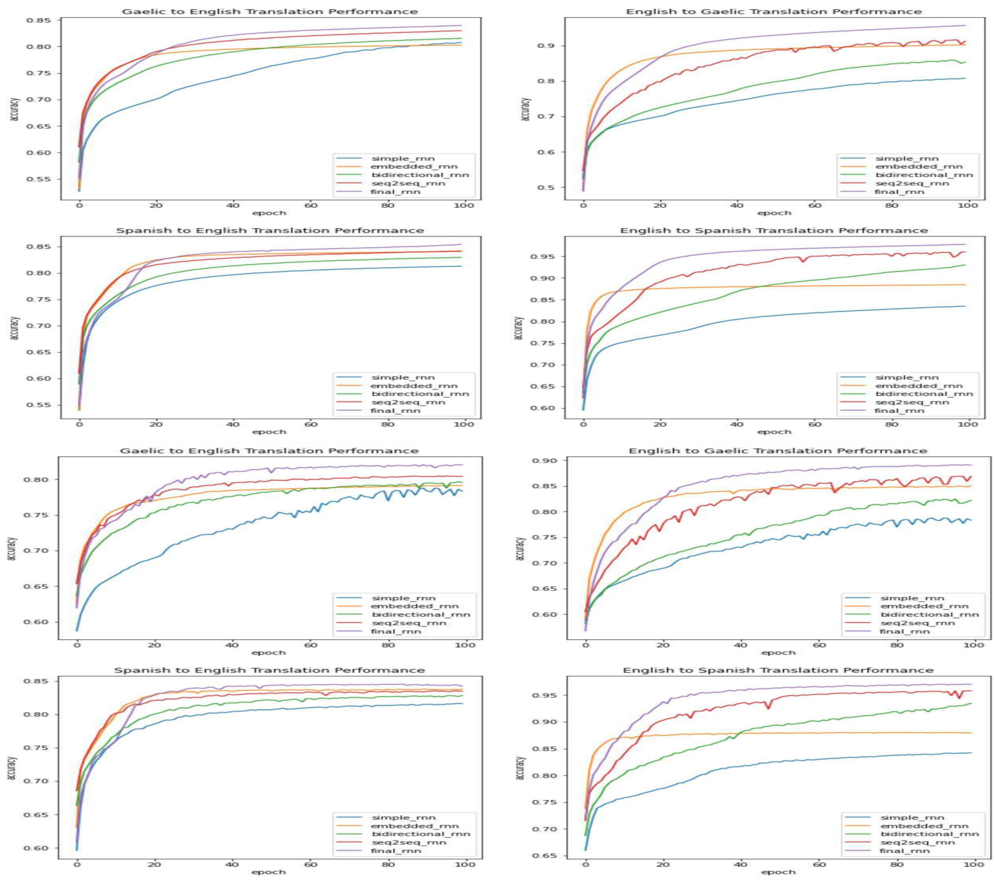


Figure 16: All RNN Model's Performance(Training and Validation)

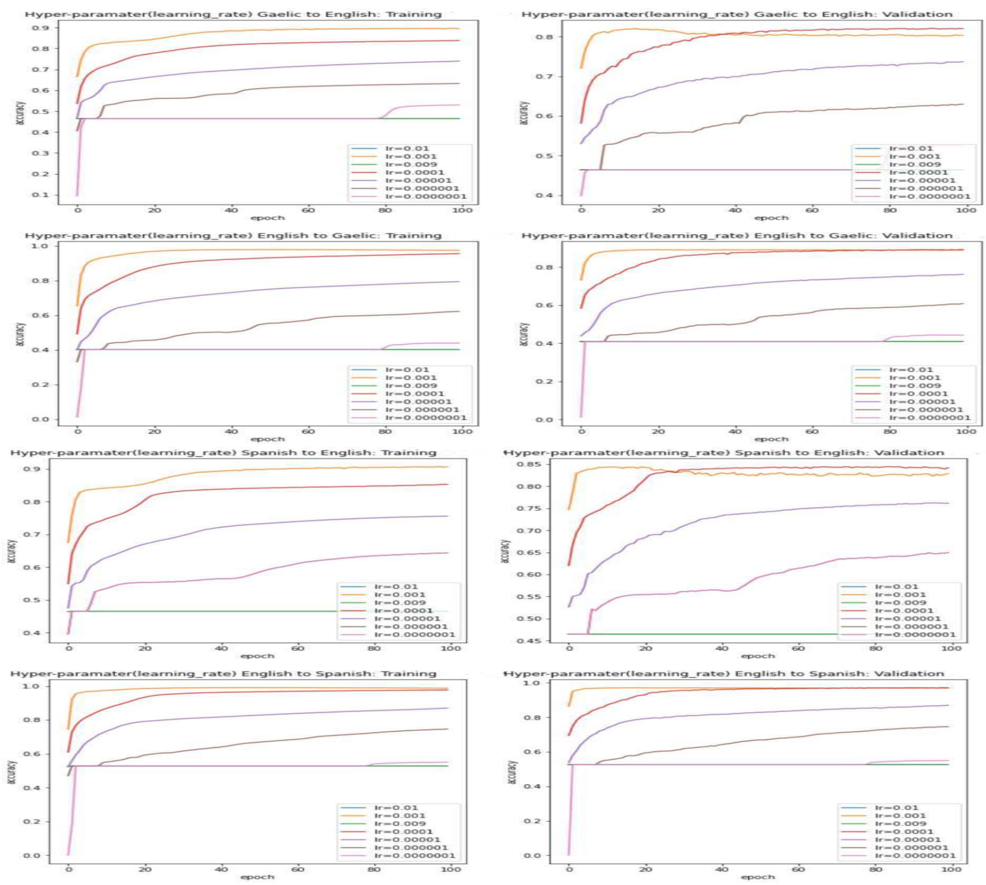


Figure 17: Hybrid RNN Model's Performance for Learning Rate Tuning

6. Discussion

In this work, five different RNN models are experimented under 2 pairs of languages in reciprocal way. A dataset from [12] is employed to perform the experiment. This dataset has been used in other studies and ML performance [19] and considered as a benchmark dataset. In total, 5 × 4 experiments are performed (5 models, 4 pairs of languages for each model). Four models out of these five models are stereotype models from literature while the last model is the main contribution of this work. As mentioned earlier, some of the layers of the hybrid RNN model are derived from other models (i.e., Bidirectional, Embedded and Seq2Seq models). As part of all experiments, a pre-processing step is carried out to make the data ready for neural networks. One of the observation from the pre-processing step is that stemmization and lemmitization techniques have very little impact on overall model performance.

Table 3 and Figures 8, 9, 10, 11, 12, 13, 14, 15, 6 and 7 show the performance details of all five models. Our proposed model (Hybrid model) has the best performance for training, validation accuracy and BLEU score. From performance point of view, the closest model to the Hybrid model is Seq2Seq. From BLEU score point of view, the second winner is Embedded model (except the English to Gaelic pair where Seq2Seq is the second winner). In all experiments and for all performances (e.g., training, validation and BLEU) simple RNN has the lowest performance.

Another interesting observation from all experiments is the behavior of languages under these models. On average, the highest validation accuracy for all models is for English to Spanish (90.8%) followed by English to Gaelic (83.8%) followed by Spanish to English (82.6%) and then at last Gaelic to English (79.2%). A similar behavior is observed for BLEU score where the BLEU score on average for all models is for English to Spanish (0.64) followed by English to Gaelic (0.60) followed by Spanish to English (0.588) and at last Gaelic to English (0.568).

The best validation accuracy from all models and all pairs of languages belongs to English to Spanish under our proposed model (97%) followed by Bidirectional and Seq2Seq models for the same pair of languages (English to Spanish) (93%). The best BLEU scores belong to Hybrid model (0.85 and 0.84 for English to Spanish and Spanish to English respectively) followed by Embedded model for English to Spanish and Spanish to English (0.80 and 0.80 respectively).

The first and obvious finding of this study shows that swapping the target and source languages has impact on the actual performance for each model. For example based on table 3 the performance metric for English to Spanish is better than Spanish to English. It shows that the language structure can be a contributing factor when training models thru RNN. Although the amount of data used for all experiment is the same, pairs of languages where Gaelic is the source language shows the worst performance. This can be an indication of differences in languages' structures.

From all the presented results here we can draw two main conclusions:

1. Hybrid model performs better than other four models on average for all quality metrics (e.g., validation and training accuracy and BLEU score). The running time per epoch for Hybrid model is the second highest (the highest one belongs to bidirectional model).

2. A quantitative comparison among these four pairs of languages indicates that English to Spanish favours the accuracy and BLEU metrics on average for all models. With almost

a large gap the second best pair of languages is English to Gaelic followed by Spanish to English and finally Gaelic to English. This indicates that pairs with English as the source language seem to have higher performance. On the other hand, among those pairs where English is the target language (i.e., Spanish to English and Gaelic to English), Spanish to English is the winner. This brings us to a new conclusion that reversing the source and target languages do not necessarily results in similar performance.

6.1. Future Work

The NLP data-sets from WMT releases [12] continue to evolve with the addition of more human languages to improve speech translation machine learning techniques, this work primarily focuses on pairs where English is either the source or target language. As a future work, other pairs of human languages e.g., Spanish to Gaelic or other European languages will be examined. We also aim to publish similar performance metric using bigger vocabulary sizes on newer WMT (e.g., WMT18 and WMT19) and other machine learning NLP data-sets.

Although the primary focus of this work is on analysis of RNN based models, as future work, more advanced structures such as transformers will be examined.

References

[1] N. Yamashita, R. Inaba, H. Kuzuoka, T. Ishida, Difficulties in establishing common ground in multiparty groups using machine translation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2009, pp. 679–688.

[2] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, S. Jain, Machine translation using deep learning: An overview, in: 2017 International Conference on Computer, Communications and Electronics (Comptelix), IEEE, 2017, pp. 162–167.

[3] T. Mikolov, M. Karafiát, L. Burget, J. Černockỳ, S. Khudanpur, Recurrent neural network based language model, in: Eleventh annual conference of the international speech communication association, 2010.

[4] I. Sutskever, O. Vinyals, Q. V. Le, Sequence to sequence learning with neural networks, in: Advances in neural information processing systems, 2014, pp. 3104–3112.

[5] D. Dong, H. Wu, W. He, D. Yu, H. Wang, Multi-task learning for multiple language translation, in: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2015, pp. 1723–1732.

[6] M. Sundermeyer, T. Alkhouli, J. Wuebker, H. Ney, Translation modeling with bidirectional recurrent neural networks, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 14–25.

[7] Y. Hu, M. Auli, Q. Gao, J. Gao, Minimum translation modeling with recurrent neural networks, in: Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics, 2014, pp. 20–29.

[8] C. Wang, M. Li, A. J. Smola, Language models with transformers, 2019.

[9] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017. arXiv:1706.03762.

[10] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, 2020. arXiv:1906.08237.

[11] Jing, Ran, Iop, A Self-attention Based LSTM Network for Text Classification, 2019, p. 012008. URL: https://iopscience.iop.org/article/10.1088/1742-6596/1207/1/012008. doi:10.1088/1742-6596/1207/1/012008.

[12] O. Bojar, C. Buck, C. Federmann, A. s. Tamchyna, Findings of the 2014 workshop on statistical machine translation, in: Proceedings of the Ninth Workshop on Statistical Machine Translation, Association for Computational Linguistics, Baltimore, Maryland, USA, 2014, pp. 12–58. URL: http://www.aclweb.org/anthology/W/W14/W14-3302.

[13] Han, SuHun, Google, googletrans: Free Google Translate API for Python. Translates totally free of charge., 2018. URL: https://github.com/ssut/py-googletrans.

[14] AiTranslate, LSI, Corportion, Languages | Ai Translate | 120+ Languages | 64 MT Language Pairs, 2020. URL: https://aitranslate.com/languages/.

[15] Z. Francois Chollet, A. G. Tan, Konduit, J. Allaire, T. O'Malley, Keras documentation: SimpleRNN layer, 2014. URL: https://keras.io/api/layers/recurrent_layers/simple_rnn/.

[16] X. Z. Geoffrey Irving, Improving the accuracy using pre-trained word embeddings on deep neural networks for Turkish text classification, volume 541, 2015, pp. 123–288. URL: http://www.sciencedirect.com/science/article/pii/S0378437119318436. doi:10.

1016/j.physa.2019.123288.

[17] Abadi, P. B. a. Ashish Agarwal, Bidirectional Layers in TensorFlow Core v2.4.0, 2015. URL: https://www.tensorflow.org/api_docs/python/tf/keras/layers/Bidirectional.

[18] H. S. Y. B. Kyunghyun Cho, Fethi Bougares, Learning phrase representations using RNN encoder-decoder for statistical machine translation, volume abs/1406.1078, 2014. URL: http://arxiv.org/abs/1406.1078. arXiv:1406.1078.

[19] MLCommons, MLCommons - Home, 2021. URL: "https://mlcommons.org/".

[20] A. Einstein, Zur Elektrodynamik bewegter Körper, volume 322, 1905, pp. 891–921.

[21] M. Goossens, F. Mittelbach, A. Samarin, The companion, Addison-Wesley, 1993.

[22] J. Tenni, A. Lehtola, C. Bounsaythip, K. Jaaranen, Machine learning of language translation rules, in: IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No. 99CH37028), volume 5, IEEE, 1999, pp. 171–177.

[23] K. Papineni, S. Roukos, T. Ward, W.-J. Zhu, BLEU: a method for automatic evaluation of machine translation, in: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Association for Computational Linguistics, Philadelphia, Pennsylvania, 2002, pp. 311–318. URL: https://doi.org/10.3115/1073083.1073135. doi:10.3115/1073083.1073135.

[24] M. o. J. EUPARL, Euparl corpus for irish-english data, in: Gaois bilingual corpus of English-Irish legislation (processed), 2014.

[25] D. Merkx, S. L. Frank, Comparing transformers and rnns on predicting human sentence processing data, 2020.

[26] D. J. Kutylowski, DeepL Translate, 2020. URL: https://www.DeepL.com/translator.

[27] M. J. Varela-Salinas, R. Burbat, et al., Google translate and deepl: breaking taboos in translator training, 2018.

[28] M. Translation, Papers with Code - Machine Translation, 2021. URL: https://paperswithcode.com/task/machine-translation.

[29] R. Karim, Illustrated: Self-Attention, 2017. URL: https://towardsdatascience.com/illustrated-self-attention-2d627e33b20a.

[30] E. Liu, Y. Chu, L. Luan, G. Li, Z. Wang, Mixing-rnn: a recommendation algorithm based on recurrent neural network, in: International Conference on Knowledge Science, Engineering and Management, Springer, 2019, pp. 109–117.