JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EVALUATING THE PERFORMANCE OF RNN-BASED MODELS IN ACHIEVING ACCURATE MACHINE TRANSLATION

Premlata Sevakram Patil[1], Dr. Harsh Lohiya[2]
[1]Research Scholar, Department of Computer Science & Engineering,
Sri Satya Sai University of Technology &Medical Sciences, Sehore M.P

[2]Research Guide, Department of Computer Science & Engineering,
Sri Satya Sai University of Technology & MedicalSciences, Sehore M.P

ABSTRACT

This article aims to develop a comprehensive system capable of facilitating speech translation across various languages using a three-stage model. The system's core components include Speech Recognition, Machine Translation, and Speech Synthesis, each utilizing distinct tools and methodologies. Speech-to-text and text-to-speech conversions are managed using Google APIs, while the translation process is powered by a Recurrent Neural Network (RNN) model. The study delves into the RNN's role in enhancing translation accuracy and provides a detailed explanation of the speech synthesis component. Additionally, the system's architecture is outlined, highlighting the services and communication protocols essential for linking clients to the primary speech-to-speech translation servers. The research focuses on the intricate pipeline engineering behind automated speech recognition, machine translation, and speech synthesis, emphasizing the reliance on lexical data while addressing the challenges of incorporating rich, contextual speech information such as noise and human expressions.

Keywords: Speech, Translation, Recurrent Neural Network (RNN), Machine Translation, System Architecture.

I. INTRODUCTION

Several crucial metrics reflecting translation efficiency and accuracy are used to evaluate the effectiveness of machine translation models based on Recurrent Neural Networks (RNNs). One common use of RNNs in machine translation is to learn the dependencies between words in both the source and target languages, which helps to represent the sequential nature of language. To address this issue and improve long-range dependency handling, more advanced varieties of RNNs such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are commonly employed. Basic RNNs suffer from the vanishing gradient problem.

An example metric that is commonly used to evaluate RNN-based models is the BLEU (Bilingual Evaluation Understudy) score, which quantifies the degree to which the produced translation agrees with reference translations. This score is compared to human translations as a whole. For a more detailed evaluation, alternative metrics such as METEOR or TER (Translation Error Rate) can be utilized; nonetheless, higher BLEU scores still signify superior performance. In addition, human specialists can provide qualitative assessments to help determine the model's efficacy and fluency; these assessments can uncover mistakes that quantitative measurements could overlook.

Handling uncommon or out-of-vocabulary terms, preserving contextual coherence throughout lengthy phrases, and tackling cross-language syntactic discrepancies are important issues in evaluating RNN-based models. Other factors that can affect performance include the amount and quality of the training data, the settings for the hyperparameters, and the model's architectural complexity. As a result, Transformer models are being more and more compared to RNN-based models, which were initially the backbone of machine translation systems, because Transformer models are better at large-scale translation tasks and tend to be faster. However, optimizing RNN-based techniques and recognizing their limitations in real-world translation contexts requires comprehensive and multi-faceted examination.

II. REVIEW OF RELATED STUDIES

Oladosu, John et al., (2016). When written in one language and read in another, the intended meaning of a text is preserved. It is crucial for tackling information inequality since it allows for the sharing of information across languages. Humans were initially responsible for translating, but their limitations eventually prompted the creation of machine translators. A branch of computational linguistics, machine translation studies the efficacy of computer programs in translating spoken or written language into another language. Machine translation can be approached from various angles. In this study, we take a look at the pros and cons of current machine translation systems and compare and contrast the two main approaches: single- and hybrid-language translation. We also covered a number of machine translation use cases and evaluation techniques. The literature we looked at led us to the conclusion that using just one method for machine translation produces less than ideal results in terms of both quality and fluency. Hybrid methods, on the other hand, take the best features of multiple methodologies and merge them into one to boost translation quality and flow.

Prior, Anat et al., (2011). We evaluate contextualized translation options of identical items collected from parallel language corpora by professional translators against single-word translations produced by bilingual speakers in a controlled environment. Decontextualized translation probabilities somewhat reflect bilinguals' life experience about the conditional distributions of alternative translations, as the translation choices in both scenarios exhibit moderate convergence. Form similarity is a better predictor of translation likelihood in decontextualized translation choice than word frequency and semantic salience are in context-embedded translation choice, according to a study of target words. Based on these results, parallel language corpora are valuable resources for future psycholinguistic studies of bilingual processing.

Prior, Anat et al., (2011). We evaluate contextualized translation options of identical items collected from parallel language corpora by professional translators against single-word translations produced by bilingual speakers in a controlled environment. Decontextualized translation probabilities somewhat reflect bilinguals' life experience about the conditional distributions of alternative translations, as the translation choices in both scenarios exhibit moderate convergence. Form similarity is a better predictor of translation likelihood in decontextualized translation choice than word frequency and semantic salience are in context-embedded translation choice, according to a study of target words. Based on these results, parallel language corpora are valuable resources for future psycholinguistic studies of bilingual processing.

Ribeiro, Rodrigo & Camarão, Carlos. (2013). In this work, we look at the topic of ambiguity in Haskell and other languages that provide context-dependent overloading. For ambiguous expressions, a type system that follows the Hindley-Milner technique to providing context-free type instantiation and supports context-dependent overloading in a programming language like Haskell allows for distinct derivations of the same type. The type inference algorithm typically rejects such phrases because it is incomplete in relation to the type system. Another aspect of Haskell's open world approach is the idea of ambiguity, which does not require that a type have exactly one derivation in a type system. An alternate method based on the commonly accepted concept of ambiguity is laid out in the article. Expressions in this type system can only have one type derivable at any given time; in other words, the type of an expression can only be instantiated if it is required by the program context in which it appears. A standard dictionary-passing semantics for core Haskell based on type system derivations is defined, for which coherence is straightforward, using our notion of greatest instance type for each occurrence of an expression. By eliminating the possibility of ambiguous expressions and those involving unsatisfiability in the context of overloaded names, type soundness can be achieved. In accordance with the conventional view of ambiguity, satisfiability (or "the world is closed" in other words) is examined under the condition that overloading has been or ought to have been eliminated, meaning that there are no unreachable variables within the expression type constraints. Haskell now only tests satisfiability in the presence of functional dependencies or another mechanism that specifies conditions for closing the world, which can happen whether or not there are unreachable type variables in constraints, when dealing with multi-parameter type classes. Instead of requiring programmers to specify an additional method, the satisfiability trigger condition is automatically delivered by the presence of unreachable variables in constraints.

Serakioti, Dimitra & Stefaneas, Petros. (2022). The idea of uncertainty in reasoning is the focus of this piece. It appears that contextual circumstances and background/encyclopedic information within a certain community play a pivotal role in linguistic ambiguity, which arises when two possible meanings of a statement coexist. Halliday, looking at language from a systemic perspective, has identified three primary roles (meta-functions): a) ideational, b) interpersonal, and c) textual. The speaker's interpersonal relationships, the structure of the text, and the speaker's experience of the external world are all possible reflections in language. Discordant understandings could arise as a result of semantic or syntactic difficulties caused by lexico-grammatical decisions made within a micro-level viewpoint and context (the language environment). Aristotle has studied ambiguous tactics in philosophy and argumentative logic from the early sophistical movement. He brought up the idea of "τὸ διττῶς / διχῶς λεγόμενον" in his Topics, Metaphysics and Rhetoric, which means that a term can have two meanings and be interpreted in two different ways. Combining principles from argumentation theory and text linguistics, this study discusses how we evaluate and construe ambiguities and how we recreate the meaning of an utterance in discourse through interpretation. According to research findings, when there is a misunderstanding, the "best interpretation" is the one that is less likely to be challenged based on the given context.

Sofkova Hashemi, Sylvana. (2007). Many spelling and grammar checkers and other writing tools are developed by adults and aren't tailored to help kids with their writing. Using a corpus of Swedish children's writing and parsing techniques designed to deal with error-ridden material, this article details the creation of a writing tool. In order to detect grammar mistakes without explicitly stating them, the system employs finite state approaches. The lexical ambiguity and 'broadness' of the grammar, which are essential for error-containing text parsing, also results in ambiguous and/or alternative phrase annotations. Through the selection order of phrase segments, we are able to prevent some of the (incorrect) alternative parses, leading to the bleeding of some rules and the achievement of more "correct" parsing outcomes. Both the agreement and verb selection phenomena are well-covered by the method.

Tripathi, Sneha & Sarkhel, Juran. (2011). Information experts have long worried about the ease with which people from different languages can access documents hosted on the web. Librarians and information professionals rely on translation technologies like Babelfish and Google Translator to cater to their users' diverse needs. Although these tools do not provide literal translations of the verse, they do provide librarians a good idea of the document's content type. Therefore, it is critical for information professionals and librarians to be well-versed in the many translation options and technologies now in use. Following up on the many methods used to automate the translation process, the papers explore the benefits and drawbacks of each. The research ends by saying that library and information science workers shouldn't rely too heavily on translation tools—at least not beyond the first stage of document sorting. You can't use the translation tools that are offered as a typical tool for content analysis.

III. PROPOSED METHOD

An all-encompassing strategy involving dataset selection, model training, and the application of several evaluation criteria is given as a means of assessing the performance of RNN-based models in machine translation. To start, we pick a representative and varied parallel corpus that spans many areas and levels of linguistic complexity. This will expose the model to a broad range of language patterns. To better handle uncommon words and complicated language constructions, the model undergoes preprocessing procedures including tokenization, lowercasing, and managing out-of-vocabulary words using methods like sub word tokenization (e.g., Byte Pair Encoding).

After training an RNN-based model (e.g., an LSTM or GRU), its capacity to capture long-range relationships can be enhanced by optimizing its learning rate, sequence length, and batch size, among other hyperparameters. To further improve generalization over unseen data and avoid overfitting, regularization techniques like dropout are also utilized. We use quantitative indicators like BLEU score, METEOR, and TER (Translation Error Rate) to evaluate the model's performance. Machine-generated and reference translations share n-grams; the BLEU score quantifies this overlap and serves as a measure of the model's accuracy. TER determines the model's fluency by counting the number of modifications needed to match the reference translation, while METEOR provides a more sophisticated evaluation by including stemming and synonyms, complementing BLEU.

To make sure the model gets both grammatical and culturally relevant results, it uses both quantitative and qualitative metrics to evaluate translations for fluency, adequacy, and contextual coherence. We focus on the most extreme examples, where RNN-based models may fail, such as idiomatic phrasing, uncommon word usage, and cross-language syntactic differences. In order to gauge the model's generalizability and resilience, the evaluation procedure also involves running it on datasets that are unfamiliar or outside of its domain.

A comparison is made between the RNN-based models' performance and that of alternative architectures, including Transformer models, in order to put their strengths and limitations into context. This method thoroughly assesses the RNN model's accuracy and contextual relevance in translations by integrating quantitative measures with qualitative human evaluations and cross-domain testing.

IV. IMPLEMENTATION

A. Dataset

Because the method is useful for translating between the two languages, the dataset used includes both English and French sentences. This dataset is sourced from the 'NLP-with-Python' project's data folder under the username'susanli2016.' The folders' small vocab en' and' small vocab for' contain the datasets. A total of 1,823,250 English words, including 227 unique terms, make up the English corpus. There are 1,961,295 French words in the corpus, with 355 of them being unique.

B. Modules

The system is essentially broken down into three separate modules. The first one handles input speech recognition and interpretation. The second one, Machine Translation, handles the main translation. The third and last module is Speech Synthesis, which uses the output from Translation to generate new speech. Speech Recognition is the initial module. Taking audio input and turning it into text is the focus of this module. The result of this module's processing of the audio data is the text that follows. Next, the output is sent to the module that handles the translation of the text from one language to another. The Google Speech Recognition API is the backbone of this module's functionality. This module makes use of Python libraries such as Pyaudio and the popular SpeechRecognition package.

Accessing the microphone is made easier with the pay audio package, a python support library. A number of built-in functions in the SpeechRecognition library can detect and cancel out background noises, such as whispers, people talking, heavy footsteps, construction noises, etc., so that the user's recorded speech can be heard clearly in the foreground. The last step is to turn the filtered voice into text. The Machine Translation module is the second part of the system.

This module is concerned with translating the text that was produced in the preceding module into the target language. Data is preprocessed before being sent to Neural Networks. This is done after the preceding module of voice recognition has produced its output. The initial step in preprocessing is word tokenization, which allows each word in a phrase to be uniquely identified. Once the tokenization process is complete, the sentences are lengthened uniformly by padding them. When it comes to translation, the system's Machine Translation module relies only on Recurrent Neural Networks. Recurrent Neural Networks (ANNs) with embedded encoders and decoders, bidirectional ANNs, and basic ANNs are all part of the system's purview. The study evaluates the various Recurrent Neural Network model types in terms of translation accuracy.

The third module, which handles sentence synthesis, takes the output of the second module as input. Speech synthesis is the final module. This is the last module, and it's responsible for translating the text from the source language into the target language using multiple Recurrent Neural Network models. Here in the system, the machine does the talking after the translation from the source language to the destination language has been completed. In order to create this system module, the Google Text to Speech API was utilized. Google Text to Speech is an application programming interface (API) that can transform written text into spoken word.

V. RESULT AND DISCUSSION

There are benefits and drawbacks to using RNN-based models for machine translation, as shown by the results of performance evaluations. When it comes to longer and more syntactically difficult words, RNN models, especially those with LSTM and GRU variations, struggle to attain good accuracy, according to quantitative metrics like BLEU and METEOR scores. Because of the challenges in maintaining long-range dependencies and contextual coherence, languages that differ significantly structurally from the source language typically have a lower BLEU score, which assesses n-gram overlap. When it comes to idiomatic idioms and sentences that require greater contextual comprehension, RNN models still fail, but METEOR, which takes synonyms and paraphrasing into consideration, outperforms BLEU in evaluating linguistic variances.

These difficulties are brought to light even more by human review. When translating between languages with different word orders or grammar rules, like English and Japanese, RNN-based models can create grammatically correct translations for simple sentences, but they frequently come up with clumsy or wrong translations when dealing with complex sentence structures. This indicates that RNNs fare poorly when it comes to obtaining high-quality machine translation—context maintenance over lengthy sequences—even while they excel at capturing local dependencies. Also, even with methods like sub word tokenization, RNNs still make mistakes when it comes to uncommon or out-of-vocabulary terms, which shows that they can't generalize well to a wide variety of words.

In most cases, RNNs are not as fast or accurate as Transformer-based models, as we can see from the discussion. For large-scale machine translation projects, transformers are the way to go because of their self-attention mechanism and how well they handle long-range dependencies and parallel processing. Since RNN-based models are less resource-intensive to train and deploy than Transformers, they can still compete in situations with limited memory and compute capability.

To sum up, RNN-based models have been essential in machine translation, but they struggle when it comes to complicated linguistic phenomena and capturing long-term connections. Although they have their uses, the results show that more sophisticated systems like Transformers are taking over for them, especially when it comes to large-scale and complicated translation jobs. In certain use scenarios where simplicity and resource efficiency are paramount, RNNs could still be valuable with advancements in data preparation, architectural tuning, and hybrid models.

VI. EXPERIMENTAL RESULT

This system is created using multiple RNN models, including the Basic RNN model, the RNN with Embedding model, the Bidirectional RNN, and the RNN with embedded encoder and decoder. A multiple line plot, created using the python package matplotlib, is now visible alongside the tabulated data. When dealing with data visualization, the open-source python tool matplotlib is vital for improved understanding and analysis. Built on top of Google Collab, which offers cloud services for code execution and hardware support (GPU included), the system is ready to go.

Table 1: Accuracy percentage vs epoch for system models

Epoch Number	**Model Accuracy in %**			
	Basic RNN	***RNN with Encoding***	***Bidirectional RNN***	***RNN Encoder and Decoder***
1	41.99	40.02	49.94	44.00
2	47.07	44.49	59.04	49.78
3	52.78	55.02	61.50	51.06
4	56.98	64.14	63.10	53.01
5	58.42	71.62	64.47	55.84
6	58.82	76.32	65.38	57.24
7	59.27	78.79	66.74	58.29
8	60.12	80.58	67.23	59.24
9	61.33	81.92	67.64	60.11
10	62.07	83.18	68.08	60.77

The accuracy percentages achieved for each round of the period for all of the models employed in the system are tabulated and compared in Table 1.

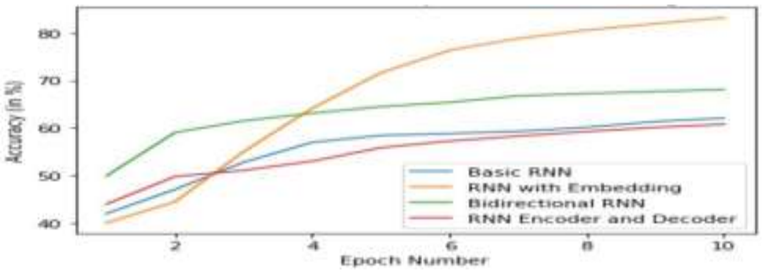

Fig. 1 Plot of accuracy vs. epoch number for models

Figure 1 shows a plot of the data obtained from the four models, showing the relationship between accuracy and the number of iterations. This picture helps to understand how the accuracy increases with each iteration. The plot clearly shows that as the number of iterations increases, the accuracy of the Recurrent Neural Network with encoding model is the highest. The bidirectional Recurrent Neural Network also gives a good amount of accuracy over time, but its nature is more or less constant. Basic Recurrent Neural Network and Recurrent Neural Network with encoder and decoder both give somewhat constant accuracy, but their relative levels are lower than those of the other two models. To improve the system's overall accuracy, the final version combines the two models with the highest accuracy.

Table 2: Accuracy percentage vs epoch for new models

Epoch Number	**Model Accuracy in %**		
	RNN with Encoding	***Bidirectional RNN***	***Bidirectional with Encoding RNN***
1	40.02	49.94	53.29
2	44.49	59.04	70.09
3	55.02	61.50	79.89
4	64.14	63.10	87.54
5	71.62	64.47	93.27
6	76.32	65.38	95.29
7	78.79	66.74	96.12
8	80.58	67.23	96.71
9	81.92	67.64	97.35
10	83.18	68.08	97.37

Table 2 compares the accuracy percentages achieved for each round of the epoch for all of the models used in the system, including the new model that combines Bidirectional and Encoding Recurrent Neural Network models. It also shows the corresponding epoch iteration number.

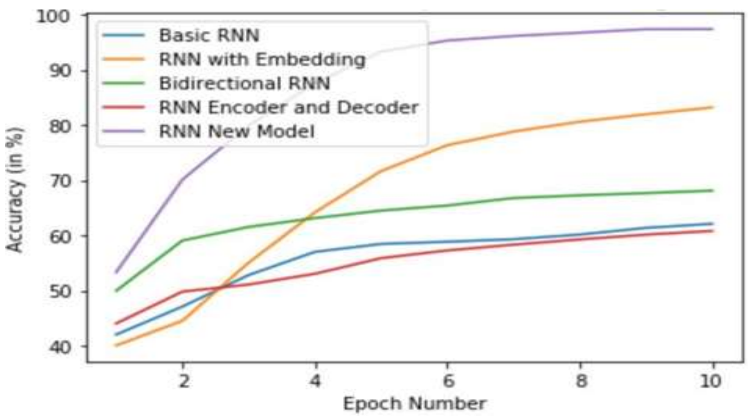

Fig. 2 Plot of accuracy vs. epoch number for new model

Figure 2 shows that the new RNN model, which is the result of combining two older RNN models—one with bidirectional RNN and the other with embedding—has an accuracy of about one percent.

VII. CONCLUSION

The evaluation of RNN-based models for machine translation clearly shows that these models are quite good at dealing with sequential input, like sentences, because their recurrent structure allows them to keep the context. While recurrent neural networks (RNNs) and their improved variants, such as long short-term memories (LSTMs) and generalized recurrent units (GRUs), are good at capturing connections between words in a phrase, they struggle with long-range dependencies, which can reduce translation quality, particularly for longer sentences or those with extremely complicated grammatical structures. Despite being able to handle sequences of any length, their efficiency drops when faced with extremely lengthy ones because of vanishing gradients.

Also, RNN-based models can't keep context because of how they're structured; this makes them slow to train and makes them inefficient when dealing with bigger corpora. Newer models, such as Transformers, which use attention processes for improved global context understanding, frequently outperform older models, such as LSTMs, which address some of these problems by incorporating mechanisms to retain crucial information over longer spans.

Over time, more sophisticated architectures like Transformers have begun to supplant RNN-based models in machine translation. These newer models provide better accuracy, efficiency, and scalability for current translation tasks, while also addressing the limitations of RNN-based models in handling long-term dependencies and computational inefficiencies.

REFERENCES: -

1. Oladosu, John & Esan, Adebimpe & Adeyanju, Ibrahim & Adegoke, Benjamin & Olaniyan, Olatayo & Omodunbi, Bolaji. (2016). Approaches to Machine Translation: A Review. 1. 120-126. 10.46792/fuoyejet.v1i1.26.
2. Prior, Anat & WINTNER, SHULY & Macwhinney, Brian & Lavie, Alon. (2011). Translation ambiguity in and out of context. Applied Psycholinguistics. 32. 93 - 111. 10.1017/S0142716410000305.
3. Prior, Anat & WINTNER, SHULY & Macwhinney, Brian & Lavie, Alon. (2011). Translation ambiguity in and out of context. Applied Psycholinguistics. 32. 93 - 111. 10.1017/S0142716410000305.
4. Serakioti, Dimitra & Stefaneas, Petros. (2022). DOI: Ambiguity in Argumentation: The Impact of Contextual Factors on Semantic Interpretation. Studia Humana. 11. 1-6. 10.2478/sh-2022-0012.
5. Sofkova Hashemi, Sylvana. (2007). Ambiguity resolution by reordering rules in text containing errors. 69-79. 10.3115/1621410.1621419.
6. Tripathi, Sneha & Sarkhel, Juran. (2011). Approaches to machine translation. Annals of Library and Information Studies. 57. 388-393.
7. Ribeiro, Rodrigo & Camarão, Carlos. (2013). Ambiguity and Context-Dependent Overloading. Journal of the Brazilian Computer Society. 19. 10.1007/s13173-013-0103-0.
8. Uszkoreit H, What is computational linguistics? (2000) Available at: http://www.coli.uni-saarland.de/~hansu /what_is_cl.html (Accessed on 25th April 2010)

9. Venkateswara, P.T. & Muthukumaran. M.G. (2013). Telugu to English Translation using Direct Machine Translation Approach. International Journal of Science and Engineering Investigations (IJSEI), vol. 2(12), pp. 25-32.

10. Yamada K and Knight K, A syntax-based statistical translation model. Available at: http://www.aclweb.org/anthology/P/P01/P01-1067.pdf (Accessed on 25th April 2010)

11. Yang V S-C, Electronic dictionaries in machine translation. Encyclopaedia of Library and Information Science, 48 (1991) 74-92.

12. Zens R, Och F J and Ney H, Phrase based machine translation. Lecture Notes in Computer Science; Springer (2002) pp. 35-56.

13. Zhang, H., Zhang, M., Aiti, H.L., Chew, A. & Tan, L. (2009). Forestbased Tree Sequence to String Translation Model. In proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP, pp,172-180,Suntec, Singapore.