Sequence to Sequence Model Performance for Education Chatbot

https://doi.org/10.3991/ijet.v14i24.12187

Kulothunkan Palasundram, Nurfadhlina Mohd Sharef (✉),
Nurul Amelina Nasharuddin, Khairul Azhar Kasmiran, Azreen Azman
Universiti Putra Malaysia, Selangor, Malaysia.
nurfadhlina@upm.edu.my

Abstract—Chatbot for education has great potential to complement human educators and education administrators. For example, it can be around the clock tutor to answer and clarify any questions from students who may have missed class. A chatbot can be implemented either by ruled based or artificial intelligence based. However, unlike the ruled-based chatbots, artificial intelligence based chatbots can learn and become smarter overtime and is more scalable and has become the popular choice for chatbot researchers recently. Recurrent Neural Network based Sequence-to-sequence (Seq2Seq) model is one of the most commonly researched model to implement artificial intelligence chatbot and has shown great progress since its introduction in 2014. However, it is still in infancy and has not been applied widely in educational chatbot development. Introduced originally for neural machine translation, the Seq2Seq model has been adapted for conversation modelling including question-answering chatbots. However, in-depth research and analysis of optimal settings of the various components of Seq2Seq model for natural answer generation problem is very limited. Additionally, there has been no experiments and analysis conducted to understand how Seq2Seq model handles variations is questions posed to it to generate correct answers. Our experiments add to the empirical evaluations on Seq2Seq literature and provides insights to these questions. Additionally, we provide insights on how a curated dataset can be developed and questions designed to train and test the performance of a Seq2Seq based question-answer model.

Keywords—Education chatbot, natural language conversation, natural answer generation, question answering, sequence to sequence learning, Seq2Seq

1 Introduction

The potential of chatbots as an online tutor to provide assistance and answers to queries and questions by students is an interesting proposition and has great potential. Question answering chatbots are intelligent systems that are able to converse with humans using natural language while providing answers . Chatbots equipped with reasoning and knowledge have the capability to scale much faster than human personnel. In this perspective, chatbots usage in education can be seen as a tool for constructivism learning approach.

The increasing use of modern technology is changing the way the teaching and learning were being conducted. Chatbot in education is not new, yet, not much research has been done on this field. Since 2006, there are a number of research focusing on the usage of chatbot in facilitating the teaching and learning in different perspectives [1]–[7]. Chatbot offers an interactive way of learning, similar to the one-to-one interaction between a student and the educator. Other than simply providing the question-answering function or sharing information among the students, chatbot can offers more benefits in serving the education purposes, including to address the problem of individual customer (e.g., student, parents, alumni) support. Chatbot's growing presence is also due to its promise in cost saving by replacing human assistants, increasing user satisfaction by speeding up response time and being available 24 hours a day, has the ability to act proactively with their users, and intelligent, in the sense that it can automatically analyse conversations [8]. According to the theory of social constructivism, knowledge can be obtained and evolved through interactions between two or more people [9]. Chatbot can assist in scaffolding academic ideas where students can obtain ideas or expand present knowledge. The students then can exchange opinions with their friends and further co-construct new knowledge.

Chatbot or AI conversational tool is a prominent instrument in a personalized learning environment, which is built to improve student interaction and collaboration. It helps the students with different learning paces absorb the knowledge according to their level, not only confined to the classroom but also in distance education [4], [7]. A personalized learning environment can be provided by the educators for each student as a result from the evolvement of the artificial intelligence. Chatbot technology is also proven as an effective tool in helping the first-year students to reduce information load as well as making them feel socially connected with their educators [5].

From the educators' perspective, chatbot can support in a lot of ways. It can be used as a mass communication tool to send messages in the form of reminders and notifications [5]. A chatbot that resides in students' mobile phones is helpful to maximize this function. Chatbot can also help in the homework and assignments-related tasks such as identifying spelling and grammatical mistakes, checking homework, assigning group work and keeping track of progress and achievements of each student [7]. Educators can assess their students' progress by analysing recorded conversations on chatbot [1].

The ultimate potential of a chatbot is a huge contrast to its current and more common usage until now: as pedagogical agents (which interact the learning content with users through human like interface) and intelligent tutoring system (which provides adaptive teaching through customised instruction and feedback to the learner). A chatbot could interact in a synchronous way and provide personalisation by reacting on individual intent so the students could actively control their learning process. In this perspective, chatbot mediated learning can be classified as a technology mediated learning (TML) [10] which is described as is described as "an environment in which the learner's interactions with learning materials (readings, assignments, exercises, etc.), peers, and/or instructors are mediated through advanced information technologies.

Chatbot can be developed based on remodeling examples of common communication using (i) defined templates and rules constructed through tools such as Amazon Lex and Google's Dialog Flow and (ii) machine learning technique such as neural network which is less laboursome. Chatbot that could answer unseen questions through approach such as similarity measure, answer deduction and response generation is regarded more intelligent. One of the frequently researched model for chatbot development is the Seq2Seq model which is based on Recurrent Neural Network (RNN) based Encoder-Decoder framework [11]. One key challenge with Seq2Seq (as in other neural network models) is that there are so many settings and hyperparameters that need to be tuned in order get a good performing working model. For example, embedding type, embedding size , hidden units size, dropout rates, neural network types are some of the settings and hyperparameters that need to set and tuned. (provide list of related works that provide experiments on this.

The primary motivation for this research is to find out and clarify a key question that we had during our initial investigation of the Seq2Seq model for natural answer generation modelling. In a typical natural question answering system, the system will be first trained on a dataset with question and answer pairs. It has been found that the Seq2Seq model is able to answer those questions well. However, in a real scenario, users may pose similar questions but in different variations. For example, the question "*What is the connection between artificial intelligence and philosophy?* " can also be rephrased as "*What is the connection between philosophy and artificial intelligence?* " and the model should be able to answer both of these questions correctly although it was trained only in one of them. A good system should be able to generate correct answers for the different variations even if they are not the exact questions originally trained on. Our key questions are, if the Seq2Seq model can handle these scenarios and how will be the performance. Table 1 lists down our intention in the form of research questions to be clarified through this experiment. The outcome of this experiment is to provide us a better idea on the configurations that we should consider for our baseline Seq2Seq model for further research.

Table 1. Research Questions

ID	Research Question
RQ1	Which embedding type performs better, word or character for education chatbot?
RQ2	How does applying dropout rate impact the performance of education chatbot?
RQ3	How well can the recurrent neural network (RNN) based Seq2Seq model handle the various categories* of education chatbot?

*Refer to section 4.1 for details

Since education-based dataset to train Seq2Seq chatbot is very scarce especially based on Malay language, the secondary motivation to perform this experiment is to find out how we can curate a dataset to train the chatbot. In this work, we investigated the comparative performance of word versus character embeddings and effects of dropout [12] on the quality of answer generated using Gated Recurrent Unit [13], a variant of Recurrent Neural Network (RNN). We investigated how Seq2Seq model handle different types of questions posed to it and if the Seq2Seq model is able to learn by itself some important relationships between the data such as synonyms. We present a preliminary experiment report with in-depth analysis on our investigation.

In summary, the main contributions of this works are as follows:

i) We provide some insights into applicable settings and optimizations that can be applied by researchers when developing a sequence to sequence (Seq2Seq) model specifically for question-answering problem. For example, we reported that dropout rate of 40% improves the answering capability of the model.
ii) We shared some examples on how a specialized dataset can be curated to train a model to enable it to learn and gain certain knowledge from the dataset and avoid depending on rules or pre-trained embeddings. This is useful for unique datasets or in our case, dataset in Malay language.
iii) We have proposed different question categories in natural language that a question-answering system should be able to provide (generate) answers to.

This paper is organized into five (5) sections. First is introduction (this section), followed by related works in section two (2) where previous experiments known to us is briefly discussed. Section three (3) explains the experiment setup which includes the dataset, evaluation questions, the models evaluated and training software. Experiment results and detailed discussion is done in section 4. We conclude this paper in section 5.

2 Related Work

One of the closest and recent work that inspired us to perform this experiment is [14]. They performed a number of experiments to compare various settings and identify some optimal settings for a SeqSeq2 model. We leveraged on their findings that bidirectional encoder is better than unidirectional, beam search of size 10 is optimal, and Bahdanau's attention [15] mechanism performed better than Luong's attention [16] mechanism by implementing these settings in our baseline model as they are not part of our research questions. However, they did not do any performance benchmarking on the effects of various dropout rates (our RQ2). And our RQ3 is very unique and novel and we could not find any other experiments to answer the research question. To the best of our knowledge, there has not been any specific developments or experiment on Seq2Seq based question-answering chatbots.

2.1 Seq2Seq Model

Figure 1 shows a sample implementation of a typical Seq2Seq model with word embeddings and attention mechanism. There are three (3) key components in the model:

i) Embedding – Embedding can be of type word or character or other forms such as bigram, trigram or even a hybrid between them. The function of the embedding layer is to convert the input into a vector of real numbers that represents the input.
ii) Encoder – Usually implemented as bidirectional encoder consisting of GRU network. The function of the encoder is to process (encode) the input embeddings which are variable length vectors and produce intermediate states which are fixed lengths vectors.

iii) Decoder – Consist of GRU network. The function of the decoder is to take the fixed length encodings produced by the encoder and generate a variable length sentence using beam search decoding [11]. Additionally, in an attention-based implementation, such as shown in Figure 1, the decoder also learns to choose which encodings should be given attention to when decoding.

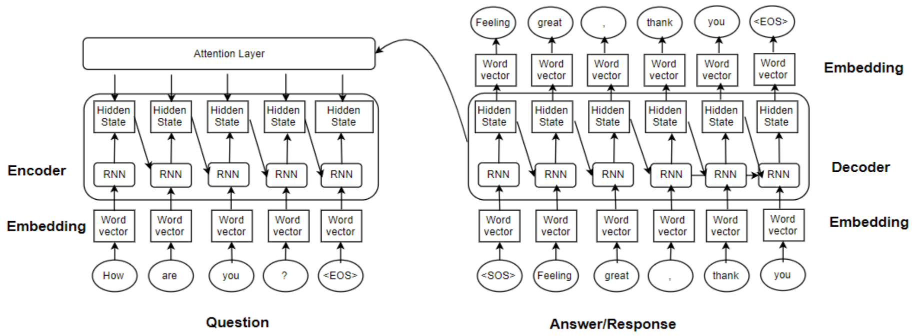

Fig. 1. Illustration of a Seq2Seq Model with Attention

As can be noted from Figure 1 above, there are many components in the Seq2Seq model, and they can be implemented differently to address certain limitation or achieve desired outcome. For example, although word embedding was originally used for Seq2Seq model, it poses a serious limitation if large dataset is used. The vocabulary will be very large and requires huge processing power and long time to train. In order to address it researchers limit the vocabulary size. However, that itself poses another limitation whereby rarely occurring words will be left out from the vocabulary thus the model learning and may impact model performance especially question-answering system. One method to address this limitation is to use character embedding [17], [18] or sub-word embedding [19]. Using character or sub-word embedding has another advantage whereby the same vocabulary can be used to continuously train the model with additional datasets. Character embedding has additional advantage over sub-word units as the vocabulary size will be very small (less than a 100) while sub-words vocabulary size may still run into tens of thousands. However, before decoding on which embedding to be used, it is wise to train and test the model to gauge the performance of word embedding versus character embedding models.

In addition to that, dropout rates should also be investigated. Dropout is a regularization technique for neural network based models proposed by [20] to reduce overfitting. When dropout is applied, randomly selected neurons are ignored during training. This forces other neurons to cover up for the missing neurons and learn more about the internal network representations.

3 Experiment Setup

3.1 Training Dataset

We manually curated an experimental tiny dataset containing one hundred (100) pairs of questions and answers (mostly in Malay language) in the domain of computer science and applications. The format of the dataset is <question><tab><answer>. Each line in the file has one question-answer pair. Table 2 shows the first five (5) question-answer pairs in the dataset.

Table 2. Part of question-answer pairs in dataset*

Line #	Text
1	Berikan contoh pembelajaran tidak diselia. Clustering.
2	Berikan contoh pembelajaran tidak diawasi. Clustering.
3	Apakah itu pembelajaran diawasi ? Dalam pembalajaran diawasi, nilai benar diberikan.
4	Apakah itu pembelajaran tidak diawasi ? Komputer akan membuat ramalannya dengan sendiri.
5	Apakah itu pembelajaran pengukuhan ? Dalam pembelajaran pengukuhan, nilai benar tidak diberikan.

The characteristics of the dataset are as described below in the following points.

i) Mapping cardinality. The dataset consists of question to answer mappings which includes one-to-one (1:1) and one-to-many (1:M) mappings of question-answer pairs. The various mappings were introduced to the dataset to find out how the variations of the Seq2Seq model handle them especially one question to many answers mapping.

ii) Synonym words. Words of same meaning are included in the dataset to see if the model can learn synonyms from the dataset. it is expected that the model is able to link these pairs (diselia, diawasi) and (neural, saraf) as synonyms.

iii) Almost identical questions or answers. Very similar questions and/or answers have been added to dataset to pose further challenge to the models to be able to really differentiate each question accurately before generating an answer.

3.2 Test Questions

We devised two (2) categories of question to test our models which are seen and unseen questions. Refer to Table 3 for details.

Table 3. Evaluation Questions

Category	Description	Count
Seen-1	One to one mapping	8
Seen-2	One to many mapping	2
Unseen-1	Different order of words	2
Unseen-2	Replaced words (in-vocabulary)	2
Unseen-3	Spelling variation (out of vocabulary)	2
Unseen-4	Additional words in question	2
Unseen-5	Less words in question	2
	Total	**20**

Seen questions are questions that the model has been trained on. Seen questions can be further divided into questions which has only one answer (one to one mapping) and questions that can have more than one correct answer (one to many mappings).

Unseen questions are questions that the model has not seen during training. Unseen questions have been devised to understand how the Seq2Seq model handles them. The unseen questions reflect reality and naturalness whereby a same question can be posed in different ways by different people. The difference includes different words with similar meaning, different order of words, a more elaborate or simpler form of question, some spelling variations due to language familiarity (or unfamiliarity) or simply typo (unintentionally). The unseen questions are divided into five (5) sub-categories as described below:

i) Words in different order – the word order of the questions is different than the seen questions
ii) Replaced words – one of the words in the seen question is replaced with another in-vocabulary word
iii) Words with spelling variations – one of the word in the seen question is spelt differently (out of vocabulary word).
iv) Additional words in question – an additional in-vocabulary word is added to a seen question
v) Shortened question sentence – one of the word is taken out from a seen question

3.3 Models Experimented

We experimented with four (4) variants of the Seq2Seq model. The models (Table 4) differ in terms of embedding types (word or character) and dropout rates.

Besides the variations as described above, all the models were configured with the following fixed characteristics and hyperparameters:

i) Bidirectional Encoder with GRU network
ii) Embedding and hidden size is 256
iii) Word embedding for answers (output)
iv) Training batch size and epoch count is 10 and 200 respectively
v) Beam search (decoding) size is 10

For the models with word embedding, the input and output vocabulary size are 177 and 245 tokens respectively. As for the character embedding model variants, the input vocabulary size is 47. They use the same output vocabulary as the word embedding model, which are 245 tokens. The token count is inclusive of software generated tokens like <unk>, <start> and <end> for both embeddings.

4 Experiment Result and Discussion

4.1 Evaluation Criteria

The models are evaluated using BLEU [21] score. BLEU stands for Bilingual Evaluation Understudy which was introduced as an automatic scoring mechanism to compare translated text with original text. In our case, we compare generated answer with gold answer. BLEU score is between 0 to 1 with 0 means no match at all and 1 means perfect match and anything in between means there are some overlaps between the two texts. For the questions which have one to many mapping either one of the answers is accepted as correct answer. Model(s) that have highest BLEU score is considered the best.

4.2 Experiment Result and Analysis

Table 4 shows the overall result for models. Bold letters indicate the best in each variant.

Table 4. Overall Model Performance (BLEU scores)

Dropout	Word Embedding	Char Embedding
0%	0.8418	0.7651
20%	0.94	0.7092
40%	**0.95**	**0.7783**
60%	0.9	0.6102

RQ1 - Which embedding type performs better, word or character for education chatbot?

Outcome. For the same dropout rates, word embedding models performed consistently better than character embedding models. The overall best word embedding model performed better than the overall best character model in terms of BLEU score, 0.95 against 0.7783.

*Analysis***.** One of the drawback of character-based models is the sequence length. For the same sentence, the sequence length for a character-based models can be several times the magnitude. This poses a challenge where character-based model may face difficulty in capturing the long-distance dependencies. This is because there are more predictions to be done by a character-based model as compared to word-based models. The more predictions to be made means there are chances to make more mistakes. In order to avoid this situation, we used character embedding only for input and word embedding for output for all variants. However, the performance of character model was still worse than word-based models.

RQ2 - How does applying dropout rate impact the performance of education chatbot?

Outcome. As shown in Table 4, applying different dropout rates does have an effect to the model performance. All the models performed best at dropout rate equals to 40%. There was slight decline in the performance of all models when no dropout or dropout rate equals to 60% was applied.

Analysis. Without dropout, the models were able to generate the correct answer for the seen questions but didn't perform as well with unseen questions, which may an indication of overfitting. When dropout was introduced, overfitting reduced but performance improved especially for the unseen questions.

RQ3 - How well can the recurrent neural network (RNN) based Seq2Seq model handle the various categories of education chatbot?

Outcome

Table 5. Model Performance On Question Category (BLEU scores)

Category	Word Embedding (40% dropout)	Character Embedding (40% dropout)
Seen-1	1	1
Seen-2	1	0.9375
Seen average	**1**	**0.9875**
Unseen-1	1	1
Unseen-2	1	0.1434
Unseen-3	1	0.5
Unseen-4	0.5	0.592
Unseen-5	1	0.6103
Unseen-average	**0.9**	**0.5691**
Overall average	0.95	0.7783

This is perhaps the crux of this experiment and make this as a novel experiment and contribution. We were interested to find out how and if the Seq2Seq models are able to handle the unseen questions and the question with one question to many answers mapping. Table 5 shows the scores for best performance for each model variant. The findings are as listed below:-

i) Word-based model was able to answer all seen questions accurately (BLEU score=1).
ii) Larger difference was found in the unseen questions category. Word-based models were able to score 0.9 while Character-based model was able to score only 0.5691.
iii) Word-based models performed well in answering unseen questions in all categories except for questions with additional words.
iv) Word model were able to learn and associate with synonym words such as diselia with diawasi and neural with saraf by correctly answering unseen question which has these words interchanged.

v) It is also interesting to note that the models generated its own version of correct answer for one of the seen questions (question with one to many mappings). The answer generated by the models were as below (mixture of diselia and diawasi in one sentence although it doesn't occur in the training dataset as shown in Table 6.

Table 6. Training and Output of Synonyms

Training Data	Berikan contoh pembelajaran tidak diselia.<tab>Clustering. Berikan contoh pembelajaran tidak diawasi.<tab>Clustering. Nyatakan kategori algoritma pembelajaran mesin ?<tab>Pembelajaran mesin diselia and tidak selia. Nyatakan kategori algoritma pembelajaran mesin ?<tab>Pembelajaran mesin di-awasi and tidak diawasi.
Question Posed	Nyatakan kategori algoritma pembelajaran mesin ?
Answer Generated	pembelajaran mesin diselia and tidak diawasi .

Analysis. This experiment has shown to us the generative power of Seq2Seq model. It was able to generate an unseen answer. Additionally, it can be noticed that the Seq2Seq model, if given the right dataset, is able to learn synonyms by itself (*no need for complementary set of rules or dictionary or pre-training of embeddings*). However, more research and experiment need to be done to make this a conclusive finding.

5 Conclusion

We provide some insights into applicable settings and optimizations that can be applied by fellow researchers when developing a Seq2Seq model specifically for question-answering problem in educational settings. Additionally, we shared some examples on how a specialized dataset can be curated to train a model to enable it to learn and gain certain knowledge from the dataset and avoid depending on rules or pre-trained embeddings which is not scalable. This is useful for unique datasets or in our case, dataset in Malay language in educational settings. We have proposed different question categories in natural language that a question-answering system should be able to provide (generate) answers to.

We conducted what we believe a novel experiment albeit being a small experiment to gain insights on few components and settings of a simple Seq2Seq model. Word embedding performed consistently better than character embedding. We demonstrated the generated power of a Seq2 Seq model and the fact that training data is crucial to get a good performing model. Although seems simple, fine tuning dropout rates showed great improvement on model performance especially in reducing overfitting and generating correct answers without the need to add any complexities to the model.

6 Acknowledgement

This work is partly the progress in a Malaysian Research University Network research grant sponsored by the Ministry of Education, Malaysia. The authors thank the funders for the kind support in the research.

7 References

[1] C. H. Lu, G. F. Chiou, M. Y. Day, C. S. Ong, and W. L. Hsu, "Using instant messaging to provide an intelligent learning environment," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4053 LNCS, pp. 575–583, 2006. https://doi.org/10.1007/11774303_57

[2] J. Jia, "CSIEC: A computer assisted English learning chatbot based on textual knowledge and reasoning," *Knowledge-Based Syst.*, vol. 22, no. 4, pp. 249–255, 2009. https://doi.org/10.1016/j.knosys.2008.09.001

[3] D. Griol and Z. Callejas, "An architecture to develop multimodal educative applications with chatbots," *Int. J. Adv. Robot. Syst.*, vol. 10, 2013.

[4] D. Song, E. Y. Oh, and M. Rice, "Interacting with a conversational agent system for educational purposes in online courses," *Proc. - 2017 10th Int. Conf. Hum. Syst. Interact. HSI 2017*, pp. 78–82, 2017. https://doi.org/10.1109/hsi.2017.8005002

[5] S. Carayannopoulos, "Using chatbots to aid transition," *Int. J. Inf. Learn. Technol.*, vol. 35, no. 2, pp. 118–129, 2018.

[6] D. Rooein, "Data-driven EDU chatbots," *Web Conf. 2019 - Companion World Wide Web Conf. WWW 2019*, pp. 46–49, 2019. https://doi.org/10.1145/3308560.3314191

[7] F. Clarizia, F. Colace, M. Lombardi, F. Pascale, and D. Santaniello, "Chatbot: An Education Support System for Student," in *CSS 2018. Lecture Notes in Computer Science*, vol. 1, Springer International Publishing, 2018, pp. 194–208. https://doi.org/10.1007/978-3-030-01689-0_23

[8] R. Winkler and M. Soellner, "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis," *Acad. Manag. Proc.*, vol. 2018, no. 1, p. 15903, 2018. https://doi.org/10.5465/ambpp.2018.15903abstract

[9] L. Vygotsky, "The Development of Higher Psychological Processes," *Mind Soc.*, vol. 6, no. 5, pp. 471–475, 1978.

[10] M. Alavi and D. E. Leidner, "Review : Knowledge Systems : Management Knowledge and Foundations Conceptual," *MIS Q.*, vol. 25, no. 1, pp. 107–136, 2001. https://doi.org/10.2307/3250961

[11] O. Vinyals and Q. Le, "A Neural Conversational Model," *Proc. 31st Int. Conf. Mach. Learn.*, vol. JMLR: W&CP, 2015.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[13] K. Cho *et al.*, "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," *Proc. 2014 Conf. Empir. Methods Nat. Lang. Process.*, pp. 1724–1734, 2014. https://doi.org/10.3115/v1/d14-1179

[14] D. Britz, A. Goldie, M. Luong, and Q. Le, "Massive Exploration of Neural Machine Translation Architectures," *Proc. ofthe 2017 Conf. Empir. Methods Nat. Lang. Process.*, 2017. https://doi.org/10.18653/v1/d17-1151

[15] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *ICLR*, pp. 1–15, 2015.
[16] M.-T. Luong, H. Pham, and C. Manning, "Effective Approaches to Attention-based Neural Machine Translation," *Proc. 2015 Conf. Empir. Methods Nat. Lang. Process.*, 2015. https://doi.org/10.18653/v1/d15-1166
[17] D. Golub and X. He, "Character-Level Question Answering with Attention," *Proc. 2016 Conf. Empir. Methods Nat. Lang. Process.*, pp. 1598–1607, 2016.
[18] M.-T. Luong and C. D. Manning, "Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models," *Proc. ofthe 54th Annu. Meet. ofthe Assoc. Comput. Linguist.*, 2016. https://doi.org/10.18653/v1/p16-1100
[19] R. Sennrich, B. Haddow, and A. Birch, "Neural Machine Translation of Rare Words with Subword Units," *ACL*, 2016. https://doi.org/10.18653/v1/p16-1162
[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.
[21] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: a method for automatic evaluation of machine translation," *Proc. 40th Annu. Meet. Assoc. Comput. Linguist.*, no. July, pp. 311–318, 2002. https://doi.org/10.3115/1073083.1073135

8 Authors

Kulothunkan Palasundram graduated from Universiti Kebangsaan Malaysia and received B.S in Computer Science (Hons) and Masters in I.T degrees in 1995 and 1998 respectively. He is currently pursuing the Ph.D. degree in Intelligent Computing in Universiti Putra Malaysia. His research interests include artificial intelligence, deep learning, big data, natural language processing and dialogue generation.

Nurfadhlina Mohd Sharef is an Associate Professor at the Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia. She got her PhD from the University of Bristol where she studied on evolving fuzzy grammar for text extraction. She is an active researcher in the field of artificial intelligence and have led projects for both fundamental and applied purposes. Her current research interests revolve around text mining, recommendation system and semantic web. Besides chatbot, her current projects are multi-objective deep reinforcement learning and technology for future learning ecosystem.

Nurul Amelina Nasharuddin received her Ph.D. from the Universiti Putra Malaysia, Malaysia in 2017. She is a senior lecturer at the Department of Multimedia, Faculty of Computer Science and Information Technology. He interests include Natural Language Processing, Cross-language Information Retrieval.

Khairul Azhar Kasmiran received his Ph.D. from the University of Sydney, Australia in 2012. He is a senior lecturer at the Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Malaysia. His interests include deep learning, reinforcement learning, performance engineering, formal verification and software development multi-task deep learning for multi-class tweets classification and deep tensor factorization model for recommendation system.

Azreen Azman is an Associate Professor at the Universiti Putra Malaysia. He received a Diploma in Software Engineering from the Institute of Telecommunication and Information Technology in 1997. Immediately, he was accepted directly to second year in Multimedia University, Malaysia to study Bachelor of Informa tion Technology majoring in Information Systems Engineering. He completed his bachelor degree in 1999. After serving in the industry for a few years, he enrolled for a PhD in January 2003, studying Computing Science specializing in Information Retrieval in the University of Glasgow, Scotland and completed his study in September 2007. His current research interests include information retrieval, text mining, natural language processing and intelligent systems. He serves as a committee member for the Malaysian Society of Information Retrieval and Knowledge Management (PECAMP) and the Malaysian Information Technology Society (MITS).

Article submitted 2019-10-05. Resubmitted 2019-11-06. Final acceptance 2019-11-06. Final version published as submitted by the authors.