Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 157 (2019) 367–374

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Computer Science and Computational Intelligence 2019 (ICCSCI), 12-13 September 2019

Indonesian Chatbot of University Admission Using a Question Answering System Based on Sequence-to-Sequence Model

Yogi Wisesa Chandra[a], Suyanto Suyanto[a,*]

[a]School of Computing, Telkom University, Jl. Telekomunikasi No. 01 Terusan Buah Batu, Bandung 40257, West Java, Indonesia

Abstract

Question and Answering (QA) system is a problem in natural language processing that can be used as the system of dialogs and chatbots. It can be used as a customer service that can provide a response to the customer quickly. A QA system receives an input in the form of sentences and produces the predictive sentences that are responses to the input. Therefore, a model that can learn such conversations is needed. This research focuses on developing a chatbot based on a sequence-to-sequence model. It is trained using a data set of conversation from a university admission. Evaluation on a small dataset obtained from the Telkom University admission on Whatsapp instant messaging application shows that the model produces a quite high BLEU score of 41.04. An attention mechanism technique using the reversed sentences improves the model to gives a higher BLEU up to 44.68.

Keywords: admission chatbot; attention mechanism; question-answering system; sequence-to-sequence

1. Introduction

A QA system is commonly used in a dialog system and a chatbot designed to handle chat like a human[1,2,3]. It can be used as a customer service to answer a question by a customer. A well-known application of the QA system for English is ALICE Bot. It is a bot chatter developed using an artificial intelligence markup language (AIML), which applies the technique of pattern recognition or pattern matching[4]. Recently, some popular QA systems have been developed for Bahasa Indonesia, such as Botika, Veronika, and AiChat Indonesia. These chatbots are commonly developed using the natural language processing (NLP) approach and the rule-based methods so that they have some drawbacks regarding flexibility and scalability.

The development of a QA system will be very difficult if it is built using a method that is a pattern matching or rule-based approach[5]. It is different from the data-driven question answering system model that can be developed

* Corresponding author. Tel.: +62-812-845-12345 ; fax: +62-022-756-4108.
E-mail address: suyanto@telkomuniversity.ac.id

10.1016/j.procs.2019.08.179

based on data or conversation history that has been carried out so that the development is enough to train the question and answering system model using existing data[6].

Some researches in deep learning show that the neural networks models indicate promising results to be used in a QA system[7,8,9]. One of them uses the sequence-to-sequence (Seq2Seq) approach that produces a good performance, such as in[10] that produces a BLEU score of 16.16 and in[11] that gives a BLEU score of 55. The BLEU score, which is a metric widely used in natural language processing (NLP)[10,12], indicates the correlation between the text generated by a machine and by a human[13]. It analyzes the frequency of n-gram in the text generated by a machine and the references provided by a human[14,15].

In this research, a sequence-to-sequence approach is combined with an attention mechanism to give a response to the given question. The attention mechanism can help sequence-to-sequence to give better results since it does not lose information contained in words that are parts of the input sentence. This combination has been done on the Ubuntu dialogue corpus and Weibo dataset with the BLEU score of 16.20[10].

The problem in this research is the need for a system that can help customer service provide a quick response to the customer and so that the customer questions are not unrequited so that questions and answering systems are made to solve these problems. This system is built using the Seq2Seq and attention mechanism approach, which is a data-driven model and then analyzed the impact when sentence input is behind the sequence. The dataset used is conversation data obtained from the Telkom University Admission or *"Saringan Masuk Bersama"* (SMB) on Whatsapp instant messaging application in Bahasa Indonesia. It consists of 2,506 training data and 397 testing data. The dataset is a pair of sentence and response (target sentence). The input to this system is the question sentence about Telkom University SMB and the output is the prediction of the answer.

This research investigates the effects of the application of attention mechanism and word order reversal in a sentence to the Seq2Seq approach in the case of question and answering system at the University of Telkom SMB to obtain a data-driven model with performance measured using the BLEU score.

A QA system is a combination of NLP and information retrieval. It is capable of answering a question automatically using a human natural language[2]. Today, QA is used in dialog systems and chatbots using a promising method called deep learning[8],[9].

A Seq2Seq is one of deep learning that is commonly used in machine translation can be adapted into a QA system[5],[9]. It is a model based on a recurrent neural network (RNN) that reads a word from an input sentence one by one and then predicts the output word, which is concatenated to be a sentence[16]. An RNN has a problem of vanishing-gradient so that some Seq2Seq models use an advanced RNN called long short-term memory (LSTM)[17]. It is frequently used to represent intelligence in a language processing[18].

In[16], the researchers get a BLEU score of 25.9 using the WMT'14 English to French dataset. Reversing words in the sentences increases the BLEU score to be 30.6 since it reduces the minimal time lag. In[11], the proposed model gives a higher BLEU Score of 55 using a data set of conversations from Twitter and Foursquare.

Another technique to improve the Seq2Seq is attention mechanism. It helps the Seq2Seq to handle the lost information of the input sentence during the encoding[19]. The encoder in attention mechanism uses a bidirectional LSTM to annotate an ordered sentence[20].

In[19], a BLEU score of 41.8 is reached for the dataset WMT'14 English to France and 28.4 for the dataset WMT'14 English to German. In[10] a BLEU score of 16.16 is achieved using Seq2Seq without attention mechanism and 16.20 for the model with attention mechanism for the Ubuntu dialogue corpus and Weibo dataset. In[21], a BLEU score of 16.92 and 37.13 are achieved using the Seq2Seq without and with attention mechanism for the dataset of Ubuntu Troubleshoot.

2. Research Method

The system developed here consists of two stages: training and testing, as illustrated in Fig. 1. The train set is a pair of input and target sentences feeding to the Seq2Seq model. The sentences are preprocessed by lowercasing, removing punctuation, and tokenizing. Next, the model is trained based on the Seq2Seq without and with attention mechanism using a learning rate of 0.001, parameter update RMSprop, and three various numbers of neurons of 100, 200, and 300. This produces a trained model that is then evaluated in the testing stage.

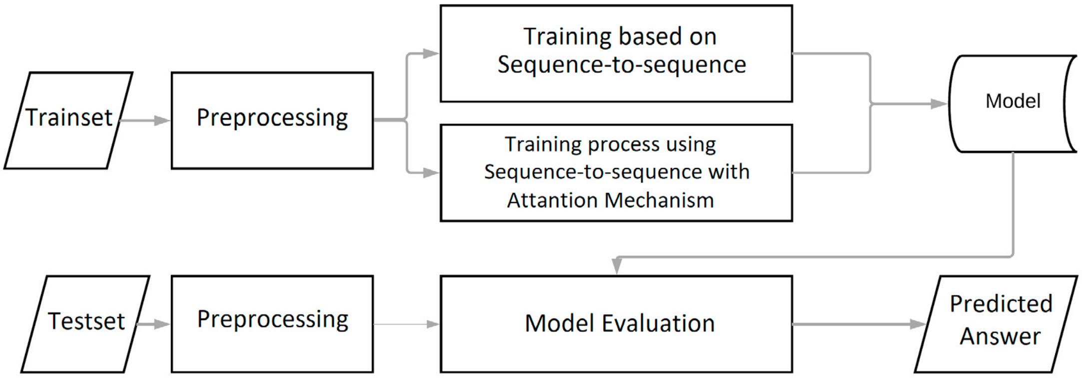

Fig. 1: Diagram of the developed system

The dataset used here is a set of Whatsapp conversations collected from the Admin of Admission in a University. A data augmentation based on synonym and typography is performed to get enough number of conversations of 2,903. It is then split into the train set of 2,506 conversations and the test set of 397 conversations. Each conversation contains one or more questions (as the input sentences) and one or more answers (as the target sentences). An example of conversation (question answering) between Admin and User is illustrated in Table 1.

Table 1: Example of the dialof of question answering between Admin and User

User	*Berapa biaya kuliah di Teknik Informatika, Telkom University?* (How much the tuition fee in Informatics Undergraduate, Telkom University?)
Admin	*Silakan kunjungi website* resmi kami di smb.telkomuniversity.ac.id (Please visit our official website smb.telkomuniversity.ac.id)
User	*Apakah mahasiswa wajib tinggal di asrama?* (Should the student stay in the dormitory?)
Admin	*Ya, mahasiswa harus tinggal di asrama untuk tahun pertama* (Yes, a student should stay in the dormitory for the first year)

The Seq2Seq model using two-layered LSTM as encoder and decoder is illustrated by Fig. 2. The former encodes the input sentence into a context vector while the later decodes the vector into the target one, which is the prediction or response to the input.

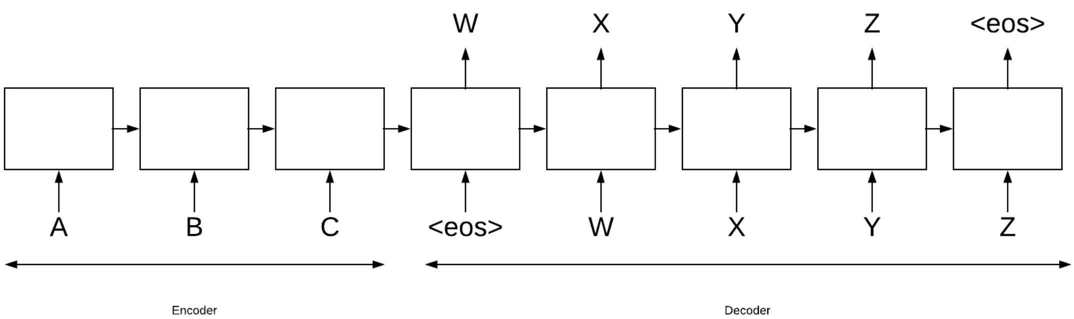

Fig. 2: Architecture of Seq2Seq model

Meanwhile, the model of Seq2Seq with attention mechanism is illustrated by Fig. 3. In this model, the encoder uses a bidirectional LSTM to annotate each word summarized into a context vector. This model takes into account both previous and next words since consists of both forward and backward LSTM. Meanwhile, the model without attention mechanism just has one context vector. The context vector depends on the annotated word (h_1, ..., h_t) mapped by the encoder from the input sentence.

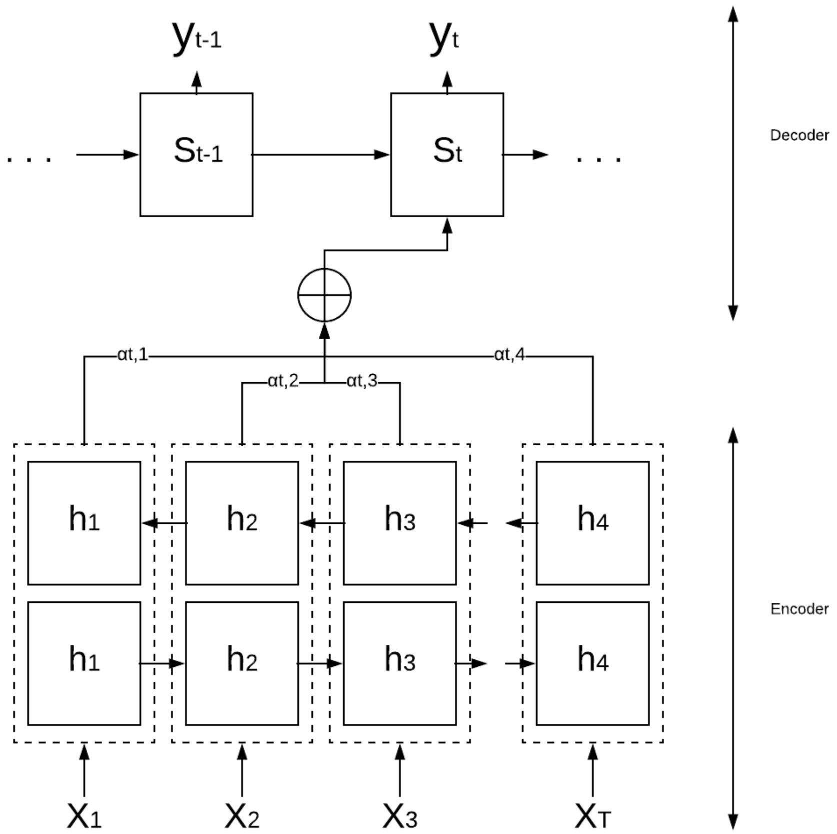

Fig. 3: Architecture of Seq2Seq model with attention mechanism

In this research, the performance of the model is measured using a BLEU score that measures the similarity between the output of the model and the human reference sentences. It is formulated as[14]

$$\text{BLEU} = \frac{\sum_{n\text{-gram}\in\hat{y}} Count_{clip}(n\text{-gram})}{\sum_{n\text{-gram}\in\hat{y}} Count(n\text{-gram})} \times 100\%, \tag{1}$$

where $Count_{clip}(n\text{-gram})$ denotes the largest number of n-grams appearing in both output and reference sentences, $Count(n\text{-gram})$ is the number of n-grams occurring in the output, and n is the length of contextual words that is set to be 4 in this research. It means that the BLEU score calculates the frequency of 4-gram in the sentence generated by the machine using the provided reference. The BLEU score is in the interval of [0, 100]. The higher the BLEU score, the more the output correlates to the human reference sentences. In most real world applications, the output sentences with the BLEU score of 30 or more is commonly considered to have a good quality those highly correlate to the human reference sentences.

3. Result and Discussion

In this research, two scenarios are performed to get the best Seq2Seq model giving the highest performance. In Scenario 1, an experiment is conducted to examine both Seq2Seq models: with and without the attention mechanism. In Scenario 2, another experiment is performed to see the impact of reversing the input sentence to the performance of both models.

3.1. Scenario 1

Fig. 4 shows that in general the more neurons the more BLEU scores. It also informs that, for all number of neurons, the model of Seq2Seq with attention mechanism produces higher BLEU scores than the model without attention mechanism. The highest BLEU score of 43.61 is achieved by the model of Seq2Seq with attention using 300 neurons. This result is caused by some context vector for each target sentence used by the model with attention mechanism keep the information of word in input sentence, which makes the model is capable of handling longer sentences.

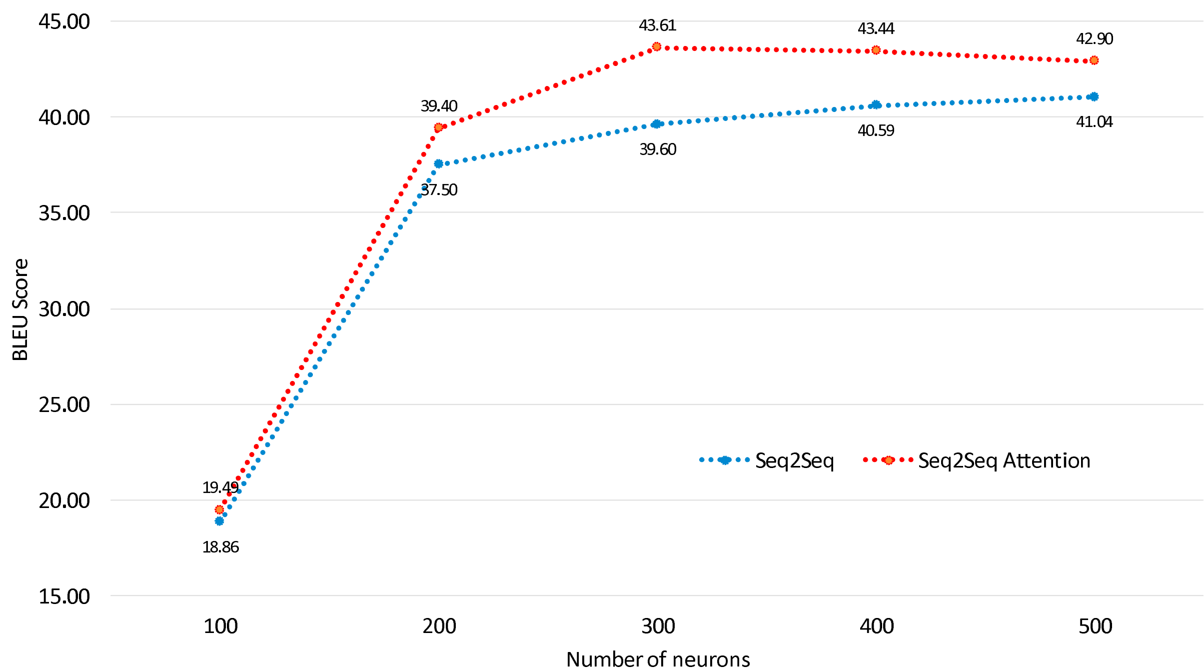

Fig. 4: BLEU Score for both models using some varying number of neurons

3.2. Scenario 2

Fig. 5 illustrates the impact of reversed sentences to the Seq2Seq model without attention. Unfortunately, the reversed sentences reduce its performance, where the BLEU score significantly decreases from 41.04 to 38.65. Reversing the sentences is expected to reduce the distance between the words at the beginning of both input and target sentences so that the problem of minimum time lag can be solved. But, further observation shows that the minimum time lag in the data set is so small that the reversed sentences are not needed.

Meanwhile, Fig. 6 shows that the reversed sentences improve the performance of the Seq2Seq with attention, where the BLEU score slightly increases from 43.61 to 44.68. This result is achieved since the encoder uses the bidirectional

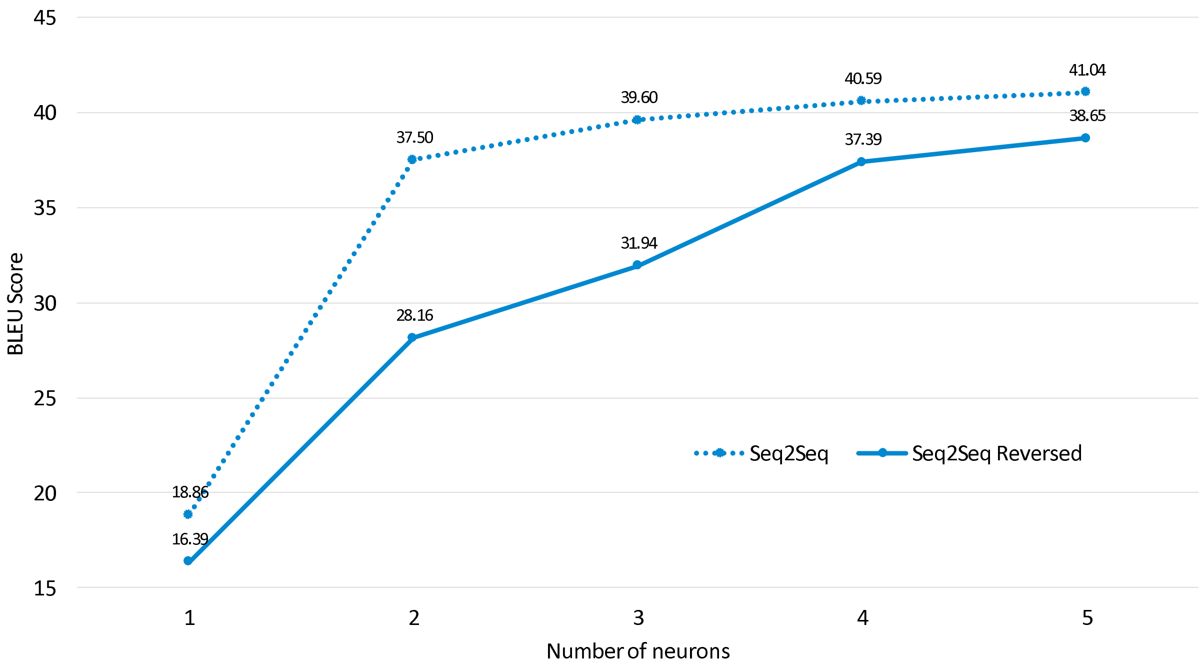

Fig. 5: Impact of the reversed input sentence in the Seq2Seq model

LSTM, which contains both forward and backward layers, so that it can annotate the input in both forward and reverse manners.

Based on the experimental results, the prototype of the Indonesian QA system is developed using the Seq2Seq with attention with the bidirectional LSTM encoder and the reversed words. The prototype can be simply described as follows. It receives a question from a user. A most similar question is then searched using the model of Seq2Seq with attention. The best answer is finally selected as the output. Table 2 illustrates the prototype of the developed Indonesian QA system.

Table 2: Examples of responses from the developed QA system prototype to the user

Question	**Response of the developed QA system**
Pagi, Min (Good morning)	*Iya, pagi juga.* (Yes, good morning)
Kak, mau tanya (I have a question)	*Ya, silakan Dik.* (Please go on.)
JPA itu apa? (What is JPA?)	*Jalur Prestasi Akademik atau JPA adalah jalur masuk melalui seleksi nilai rapor dari nilai semester.* (The Track of Academic Achievement or TAA is an admission track using academic report and grade point semester.)

In the future, this prototype can be enhanced using the subword features, instead of a word, to solve the out of vocabulary unknown words. The subword should be used is varying for different languages. For English, the commonly used subword is the character *n*-gram. Some other languages use a syllable or a morpheme as the subword. Since Bahasa Indonesia is a syllable-rich language[22], the syllable is promising to be used as the subword feature. Furthermore, this idea will be interesting since a high performance Indonesian orthographic syllabification has been opportunely developed in[23].

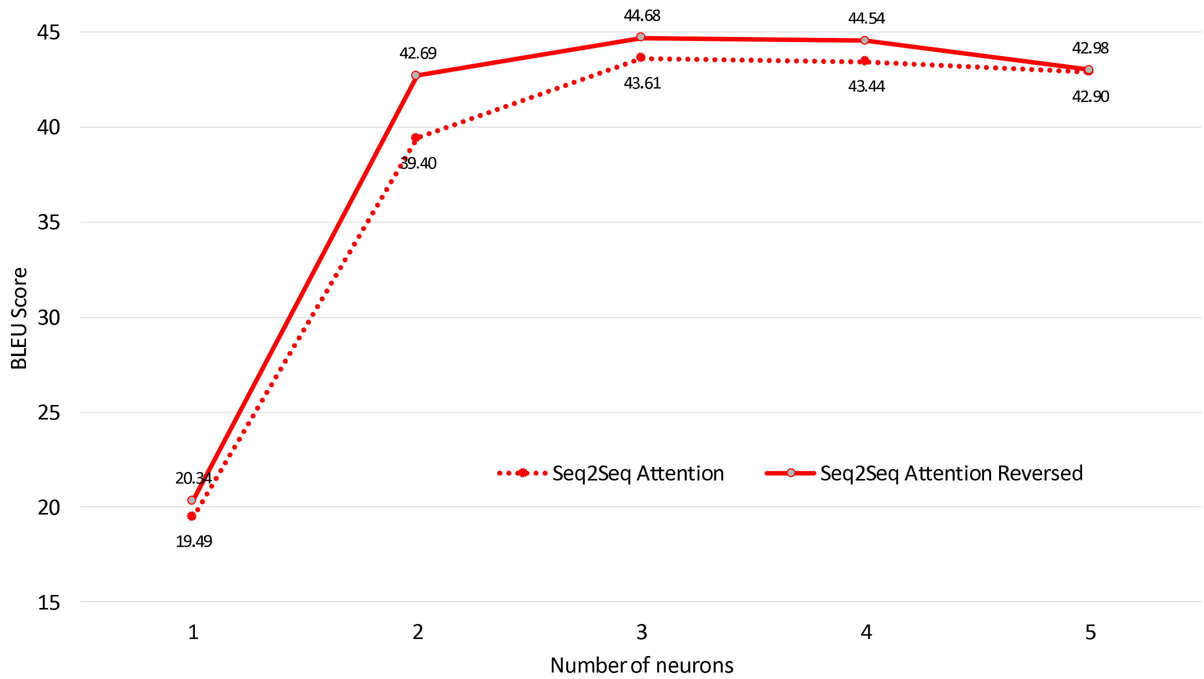

Fig. 6: Impact of the reversed input sentence in the Seq2Seq model with attention mechanism

4. Conclusion

An Indonesian QA system using a Seq2Seq approach is successfully developed. Evaluation on a small dataset from the Admission of Telkom University shows that the standard Seq2Seq model without attention mechanism gives a BLEU Score of 43.61. Both attention mechanism and reversed sentences slightly improve the model to produce a BLEU Score of 44.68, where it is achieved with an encoder of bidirectional LSTM that contains both forward and backward layers by exploiting 300 neurons. These two layers are capable of annotating the input in both forward and reverse manners.

Acknowledgment

We would like to thank all colleagues in the Admission of Telkom University for the dataset.

References

1. Deryugina, O.V.. Chatterbots. *Nauchno-Technicheskaya Informatsiya* 2010;.
2. Divyanshu Bhardwaj Partha Pakray, J.B.S.S.. Question answering system for frequently asked questions. *CEUR Workshop Proceedings* 2016; **1749**.
3. Yashvardhan Sharma, S.G.. Deep learning approaches for question answering system. *International Conference on Computational Intelligence and Data Science* 2019;.
4. Maria das Graas Bruno Marietto, R.v.d.A.. Artificial intelligence markup language: A brief tutorial. *arxiv* 2013;.
5. Lifeng Shang Zhengdong Lu, H.L.. Neural responding machine for short-text conversation. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* 2015;.
6. Yi Luan Yangfeng Ji, M.O.. Lstm based conversation models. *arXiv* 2016;.
7. Jun Yin Xin Jiang, Z.L.L.S.H.L.X.L.. Neural generative question answering. *IJCAI-16* 2016;.
8. Yan, R.. ¢chitty-chitty-chat bot¢: Deep learning for conversational ai. *IJCAI-18* 2018;.

9. Oriol Vinyals, Q.V.L.. A neural conversational model. *Proceedings of the 31 st International Conference on Machine Learning* 2015;.
10. Hainan Zhang Yanyan Lan, J.G.J.X.X.C.. Tailored sequence to sequence models to different conversation scenarios. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* 2018;:1479–1488.
11. Marjan Ghazvininejad Chris Brockett, M.W.C.B.D.J.G.W.t.Y.M.G.. A knowledge-grounded neural conversation model. *AAAI* 2018;.
12. Lantao Yu Weinan Zhang, J.W.Y.Y.. Seqgan: Sequence generative adversarial nets with policy gradient. *Association for the Advancement of Artificial Intelligence* 2017;.
13. Michel Galley1 Chris Brockett, A.S.Y.J.M.A.C.Q.M.M.J.G.B.D.. ¢bleu: A discriminative metric for generation tasks with intrinsically diverse targets. *ACL* 2015;.
14. Kishore Papineni Salim Roukos, T.W.W.J.Z.. Bleu: a method for automatic evaluation of machine translation. *Computational Linguistics (ACL)* 2002;:311–318.
15. Ryan Lowe Michael Noseworthy, I.V.S.N.A.G.Y.B.J.P.. Towards an automatic turing test: Learning to evaluate dialogue responses. *Nauchno-Technicheskaya Informatsiya* 2010;.
16. Ilya Sutskever Oriol Vinyals, Q.V.L.. Sequence to sequence learning with neural networks. *Neural Information Processing Systems Proceedings* 2014;.
17. Sepp Hochreiter, J.S.. Long short-term memory. *Neural Computation* 1997;.
18. Chenfei Wu Jinlai Liu, X.W.X.D.. Chain of reasoning for visual question answering. *NIPS* 2018;.
19. Dzmitry Bahdanau KyungHyun Cho, Y.B.. Neural machine translation by jointly learning to align and translate. *ICLR* 2015;.
20. Mathias Berglund Tapani Raiko, M.H.A.V.J.K.. Bidirectional recurrent neural networks as generative models. *NIPS* 2015;.
21. Hongyuan Mei Mohit Bansal, M.R.W.. Coherent dialogue with attention-based language models. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence* 2017;.
22. Suyanto, S., Hartati, S., Harjoko, A., Compernolle, D.V.. Indonesian syllabification using a pseudo nearest neighbour rule and phonotactic knowledge. *Speech Communication* 2016;**85**:109–118. doi:\bibinfo{doi}{10.1016/j.specom.2016.10.009}. URL http://dx.doi.org/10.1016/j.specom.2016.10.009.
23. Parande, E.A., Suyanto, S.. Indonesian graphemic syllabification using a nearest neighbour classifier and recovery procedure. *International Journal of Speech Technology* 2019;**22**(1):13–20. doi:\bibinfo{doi}{10.1007/s10772-018-09569-3}. URL https://doi.org/10.1007/s10772-018-09569-3.