INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 3)
Available online at: www.ijariit.com

Neural network conversational model for small dataset with sequence to sequence model

Mohd Nazeeb Khan
najeebagra@gmail.com
Faculty of Engineering and Technology, Agra, Uttar Pradesh

Anurag Yadav
anuragyadav13@rediffmail.com
Faculty of Engineering and Technology, Agra, Uttar Pradesh

ABSTRACT

This paper we purpose conversational model for specific domain which generates new text for two-way conservation. Conversational modeling is an important task in machine intelligence and natural language. In this paper we used a recently proposed sequence to sequence model for predicting the next sentence and the previous sentence or sentences in a conversation. Conservational model is able to extract knowledge from specific domain and noisy dataset. Main dataset subtitles. Our domain for a conservational model is movie conversation dataset. The open-domain movies transcript are noisy. The expert model can perform simple forms of common sense reasoning. Our conversational model can be end-to-end trained and required changeable rules. We expected our model pass the lack of consistency in our model and conversation process have been successfully applied in sequence- to-sequence mapping task.

Keywords: *Conversational model, Neural network*

1. INTRODUCTION

In this paper, we consider the problem of building Conservational Agent That has the ability to interact in one-on-one conversation on a Specific topic [1,7,9]. Our Paper is Focus on Small dataset Conversational Model Which Not Required very large dataset. We hypothesize that this is due to the lack of sufficiently large datasets. Our paper, we used corpus Movie dataset. Which have subtitle English Movies Conversation [14] we address the challenge of consistency and how to endow data-driven systems with the coherent needed to model human-like Conversation [20,23]. We will define an agent as the character that an artificial agent performs during conversational interactions. An agent can be viewed as a composite of elements of identity language behavior, and interaction style [21]. An Agent is also adaptive, since an agent may need to present different facets to different human interlocutors depending on the interaction.

The Conservational Model integrates a vector representation into the target part of the Sequence to Sequence Model [13]. The model encodes the interaction patterns of two interlocutors by constructing an interaction representation from their individual embedding and incorporating it into the sequence to sequence model [8]. These Agent vectors are trained on human-human conversation data and used at test time to generate personalized responses. Our experiments on a movie corpus conversations [2, 3]. This leveraging Agent vectors can improve relative performance as a communication intelligence gain in consistency as human Intelligence.

2. RELATED WORK

Our conversational Model inspires the work done by Ritter et al (2011). Ritter first Proposed Conversational Model on Micro-Blogging Site where Model Post Microblog with generative Probabilistic Approach. Generating responses was found to be considerably more difficult than translating between Languages. He views the response generation problem as a translation problem, where a post needs to be translated.

In Our Research Banchs et al (2012) [1]. The first to suggest using movie scripts to build dialogue systems. Conditioned on one or more observation, a model searches in a database of movie scripts and retrieves a sequestrate response. This was later followed up by Ameixa et al. (2014), who manifest that movie subtitles could be used to provide responses to out-of-domain questions using an information retrieval system.

3. MODEL

Our Work on Conservational model Focus on Small Datasets and Implementing Recurrent Neural Network Approach with Sequence to Sequence Framework [13, 2]. We divided our Conservational Model into parts for Better Explains. We using the response generation task as input, for conservational Agent [15, 11]. Our input sequence is "*how are you*". Each word from the input sequence is associated to a vector let M denote the input word sequence conservation$M = \{m\,1\,, m\,2\,, \ldots, m\,I\,\}$. R Denotes the word sequence in response to , where $R = \{r\,1\,, r\,2\,, \ldots, r\,J\,, EOS\}$, and J is the length of the response. Token word is associated with a K dimensional distinct word embedding. V is the vocabulary size.

3.1. The sequence to sequence Model

We use the sequence to sequence framework for conservational Model. The Sequence to Sequence Model relies on the encoder-decoder paradigm. The encoder encodes the input sequence, and the decoder generates the target sequence [19,13]. The relative this type of model has been refined over the past few years and achieved a great result. With attention. Attention is a mechanism that intensity the model to learn to focus on specific previous of the input sequence [24]. When decoding, instead of relying only on the hidden vector of the decoder in long short-term memory [5,9]. This technique is of performing attention is explained by Bahdanau et al [1]. We slightly modify the formula that we defined above by adding a new vector c_t to the input of the LSTM.

$$h_t = LSTM\left(h_{t-1}, \left[w_{i\,t-1}, c_t\right]\right)$$
$$s_t = g(h_t)$$
$$p_t = softmax(s_t)$$
$$i_t = argmax(p_t)$$

The vector $\underline{c_t}$ is the attention vector. We enumerate a new context vector at each decoding step. First, with a function $f(h_t - 1, e_{t'}) \mapsto a_{t'} \in R$, enumerate a score for each hidden state $\underline{e_{t'}}$ of the encoder [16,5,12]. After we normalize the sequence of $a_{t'}$ using a softmax and compute $c_{t'}$ as the weighted median of the $e_{t'}$.

$$a_{t'} = f(h_{t-1}, e_{t'}) \in \qquad for\ all\ t'$$
$$\overline{a} = softmax(\alpha)$$
$$c_t = \sum_{t'=0}^{n} \overline{a}_{t'} e_{t'}$$

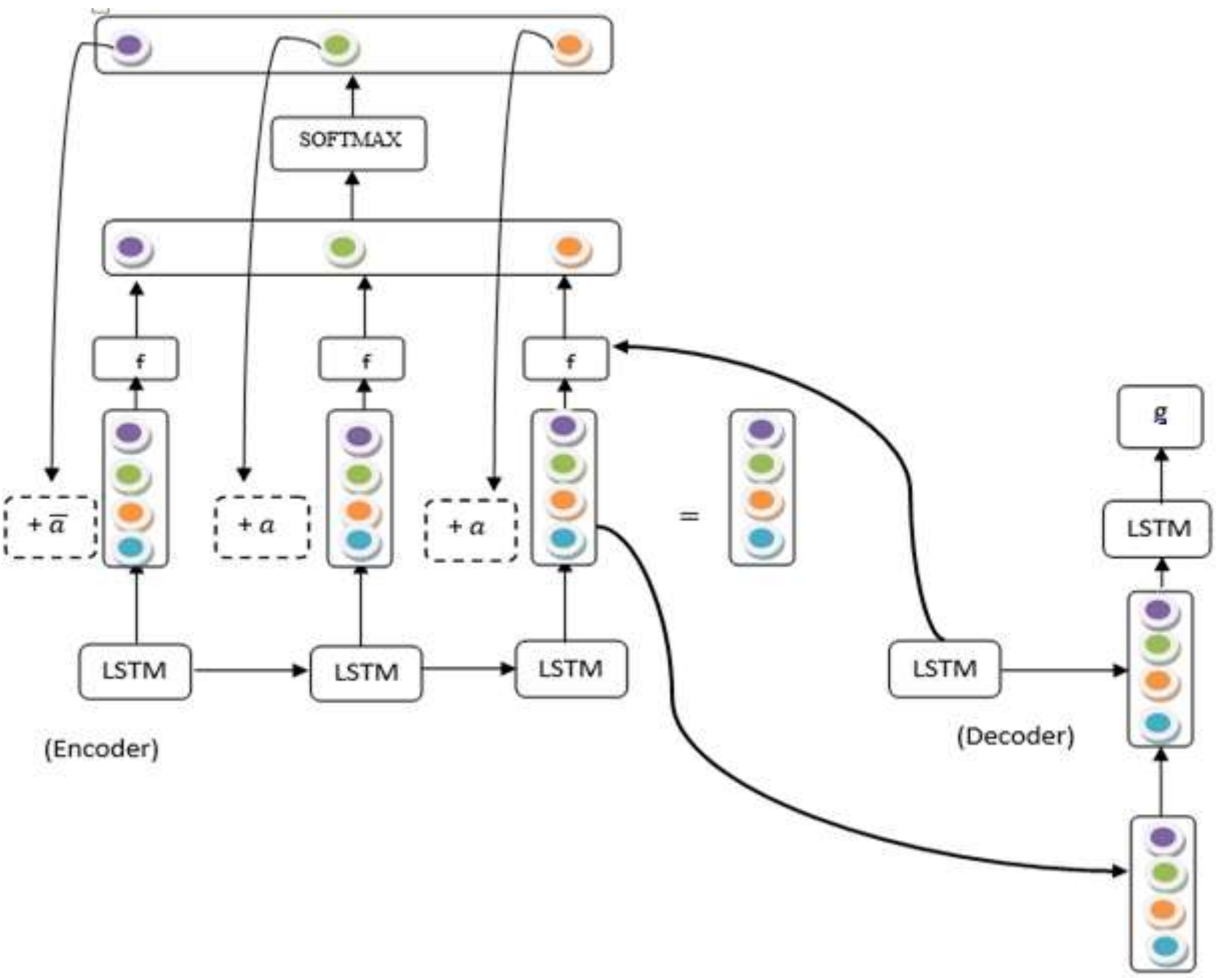


Fig. 1: Attention mechanism

The choice of the function f varies, but is generally one of the following

$$f(h_{t-1}, e_{t'}) = \begin{cases} h_{t-1}^T e_{t'} & dot \\ h_{t-1W_{et}}^T & general \\ v^t tan\,h\left(W\left[h_{t-1,} e_{t'}\right]\right) & concat \end{cases}$$

We turn out that the attention weighs $\overline{a}$ can be easily interpreted model generating the word [7, 22]. We expect $\overline{a}$ are to be close to 1 while $\overline{a}$ how and$\overline{a}$. You to be close to 0. Intuitively, the context vector . In our research sequence to sequence model will be roughly equal to the hidden vector of are and it will help to generate the English text "fine" as an input. By pushing the attention weights into a matrix. We would have access to the alignment between the input and output from the conservation.

3.2. Training

For our conservational model training, we use the predicted token as input to the next step during training errors would stockpile and the model would rarely be exposed to the correct distribution of inputs, making training slow or impossible. To enhance our training, one machination is to feed the genuine output sequence comment <sos> into the decoder's LSTM and predict the next token at every position comment <eos>.

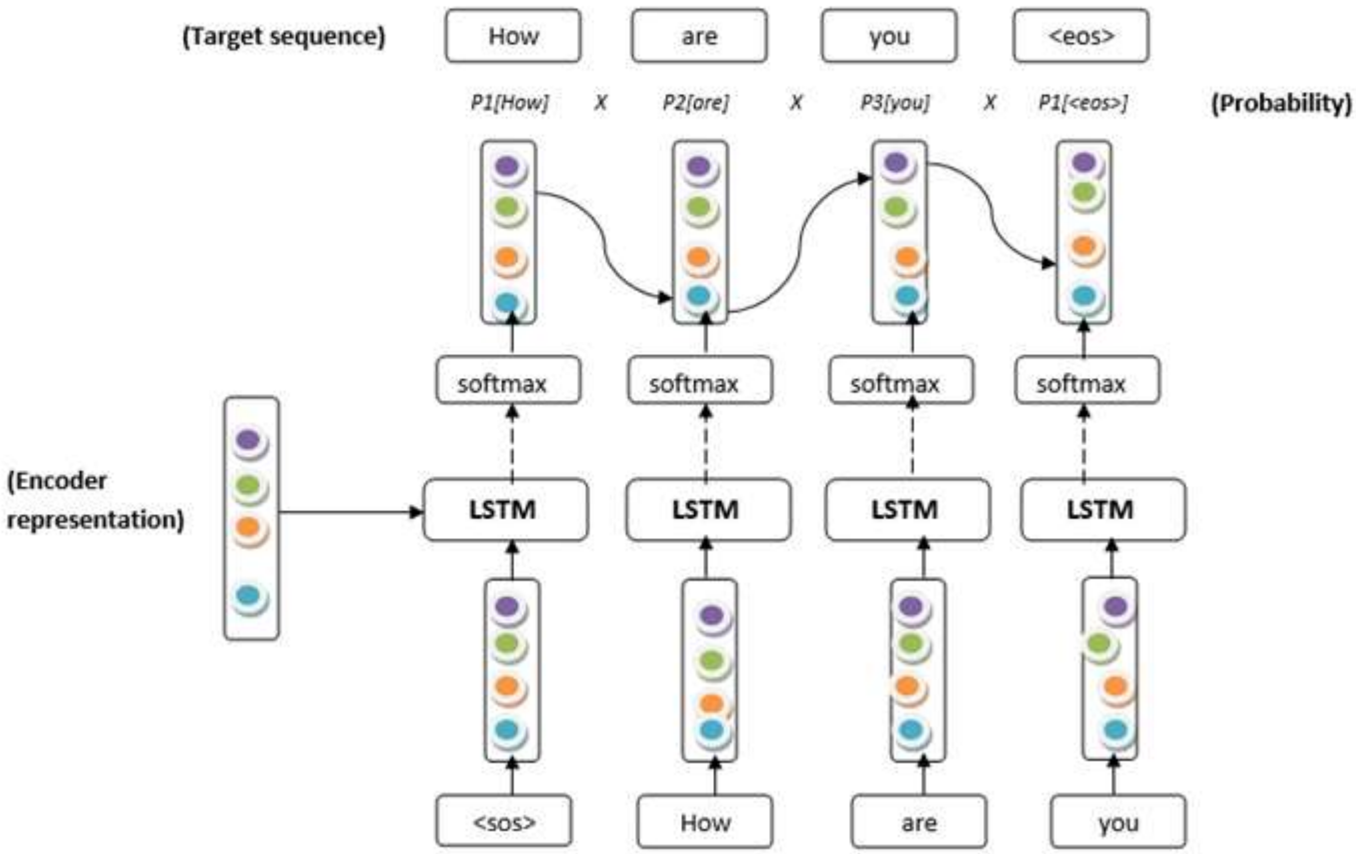
The decoder outputs vectors of probability over the vocabulary $p_i \in R^v$ for every time step. Then, for a given target sequence $\boldsymbol{y_1}, \dots, \boldsymbol{y_n}$, we can enumerate its probability as the product of the probabilities of each token [4]. In our research we see being produced at each relevant time step.

$$P(y_1, \dots, y_m) = \prod_{i=0}^{m} p_i[y_i]$$

Where $p_i[y_i]$ means that we extract the y_i entry of the probability vector pi from the i decoding step. In specific, we can enumerate the probability of the genuine target sequence [8]. A flawless system would give a probability of 1 to this target sequence, so we are going to train our network to maximize the probability of the target sequence, which is the same as minimizing.

$$-\log P\,(y_1, \dots, y_m) = -log \prod_{i=1}^{m} p_i[y_i]$$

$$= \sum_{i=1}^{n} \log p_i[y_i]$$



In our research, we actually are minimizing the cross-entropy between the target distribution and the predicted distribution outputted by our model.

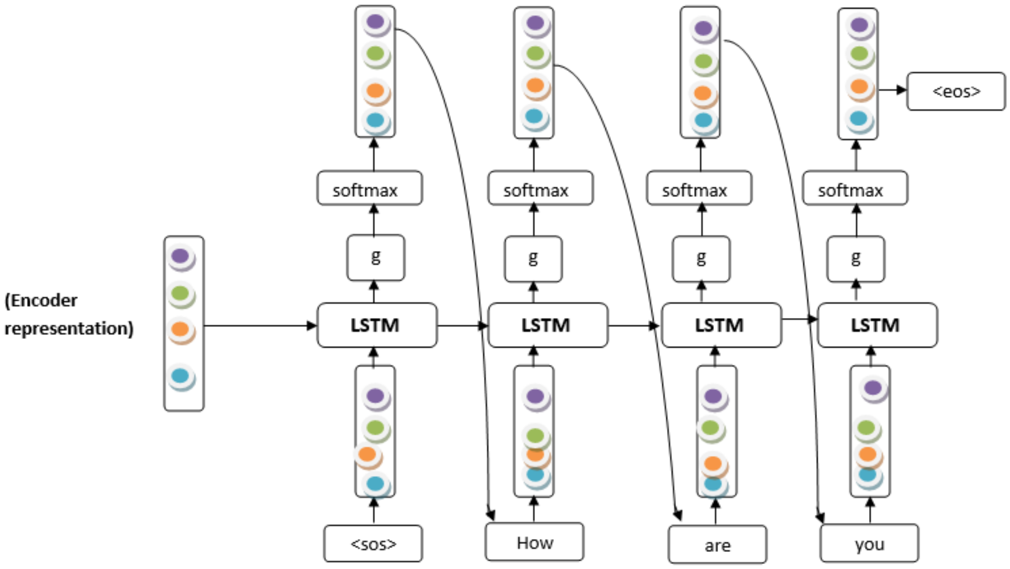
$$-logp_1[How] - logp_2[are] - logp_3[you] - logp_4[< eos >\}$$

3.3. Decoding

We find in our research conservational model take enormous time to train. We can define different behaviors. In particular, we defined a specific behavior that speeds up training. Then we approaching decoding output sequence [16]. There indeed are two main ways of performing decoding at training time the first of these methods is greedy decoding. It is the most natural way and it includes in feeding to the next step the most likely word predicted at the last step. Greedy Decoder feeds the best token to the next step in LSTM.

In our research we examine after having trained the model, it can happen that the model makes a small. But this would destroy the entire decoding. We using more reliable approach is a better way of performing decoding, in our small dataset conservational model we used Beam Search as decoding [17]. Instead of only predicting the token with the best score, we keep the course of k hypotheses. At each new time step, for these five hypotheses we have V new possible tokens. It makes a total of $5V$ new hypotheses. Then, only keep the five best ones. Formally, define H_t the set of hypotheses decoded at time step t.

$$H_t \coloneqq \{(w_1^1, \dots, w_t^1), \dots, (w_1^k, \dots, w_t^k)\}$$



For instance, if $k = 2$, one possible H_2 would be

$$H_2 \coloneqq \{(comment\ are), (comment\ you)\}$$

For our conservational model we consider all the, possible candidates C_{t+1}, produced fromH_t by adding all possible new tokens and keep the k highest scores.

$$C_{t+1} \coloneqq \bigcup_{i=1}^{k} \{(w_1^i, \dots, w_t^i, 1), \dots, (w_1^i, \dots, w_t^i, V)\}$$

$$C3 = \{(comment\ are\ comment), (comment\ are\ are), (comment\ are\ you)\}$$
$$\cup \{(comment\ you\ comment), (comment\ you\ are), (comment\ you\ you)\}$$

and for instance we can imagine that the two best ones would be.

$$H_3 \coloneqq \{(comment\ are\ you), (comment\ you\ are)\}$$

Once every hypothesis reached the <eos> token, we return the hypothesis with the highest score.

4. DATASET

In our Conservational Model, we used MovieTriples dataset for Training our Model and expanding and preprocessing dataset by Banchs et al. (2012) to make it fit the generative dialogue for our model. The dataset is available upon request. Movie transcripts span a large range of topics, contain long conservation with few participants and relatively spelling error and few acronyms. we believe bootstrapping a subject-oriented spoken dialogue system based on movie transcripts will enhance performance of our Conservational Model [11,6].

Table 1: Statistics of the dataset

	Training	Validation	Test
Movies	484	65	65
Triples	196308	24,717	24,271
Avg. tokens	53	53	55

In Our Research, we used the Python library called natural language toolkit NLTK for tokenization and name and entity recognition. With NLTK we replaced names and numbers to the <person> and <number> tokens. To minimize data sparsely further, we transformed all tokens to lowercase letters, and all but the 10,000 most chronic tokens were replaced with a generic <unk> token [11]. To capture the interactive dialogue structure, a special end-of-utterance token is appended to all comment and a continued comment token between breaks in lines from the same speaker [15]. To avoid dependencies between triples coming from another movie, we first split the movies into training, validation and test set, and then construct the triples. Reported are Statistics in above Table 1.

5. EVALUATION

In our Research, we Evaluation conservational model, with describe the experimental results with the datasets [2]. We present a rather impartial measure which is the perplexity in our model on the experimental test data. We show some samples of the conservation with the system that we trained given the lack of data for Model domain specific and clean dataset, we trained our model with small data in sequence to sequence framework [8]. We built a vocabulary of the most frequent 20K words and introduced special tokens indicating a turn to conservation.

Table 2: The generated response from the trained model.

Human	***Hi***
C.A	*Hi, there.*
Human	*my computer is not responding*
C.A	*Are you going to contact me?*
Human	*I updated my computer*
C.A	*I mean it is a good thing to do in that show!*
Human	*no, it is on Linux*
C.A	*Well, I know it is something.*
Human	*what is your name? my name is Najeeb*
C.A	*I am C.A*
Human	*did you know about India*
C.A	*No, and it was a bit of Pakistan.*
Human	*that is alright*
C.A	*It's fine. The whole motherboard is fine.*
Human	*What is your age?*
C.A	*I am 9 now.*
Human	*Do you love me?*
C.A	*Love now, love me.*
Human	*What is your major?*
C.A	*Major, I have the scholarship to say.*
Human	*see you later*
C.A	*See you again soon.*
Human	*I ' ll need to verify who it is at that machine*
C.A	*No, op is not the symptoms of being asked!*
Human	*hi, I forgot my password*
C.A	*They aren't ready for your vacation.*
Human	*What did you study in college?*
C.A	*Honestly, it was during college but they were specifically sharing.*
Human	*In which country do you live now?*
C.A	*I live in the united states.*
Human	*what is that?*
C.A	*It is a little instrument.*
Human	*we have Linux on your machine.*
C.A	*But I don't know a reason.*
Human	*I am using Linux.*
C.A	*Is artificial, but I like that time?*
Human	*what is moral?*
C.A	*Dennis side starts with a good body.*
Human	*seriously, what is morality?*
C.A	*Morality is the differentiation of intentions, decisions, and actions if you can't proper living there.*
Human	*what is the definition of living?*
C.A	*Every man, I talk like progress.*
Human	*You're the one who's crazy.*
C.A	*One now we can do that before reading.*

At convergence, our Conservational model achieved a perplexity with experimental results. above is a sample of a simulated troubleshooting session, in which Machine is our Neural Conversational Model, and Human the interacting with a Conservational agent.

6. CONCLUSIONS

In our research, we introduce a small dataset training method for conservational our model. The results presented in this paper. That indicate the training method can achieve significant performance [18]. Moreover, the training also yields a trained discriminator that can be used to select the high probability answer, when conservation different domain area which not part of our dataset. Model performance of was measured using perplexity [14]. While this is an established measure for conservational models, in the dialogue setting utterances may be submerged by many common words arising from colloquial or informal exchanges. It may be useful to research other measures of performance for generative conservation systems. We also considered actual responses produced by the model.

In research, we achieved avant-garde results on the utterance ranking problem recently introduced in [1]. The best performing model is a sequence to sequence model in neural networks. In the future, we plan to use our system as a base for more tortuous models with (GRU) gated-recurrent-unit the standard neural network paradigm.

7. REFERENCES

[1] Rafael E Banchs and Haizhou Li, (2012), “IRIS: a chat-oriented dialogue system based on the vector space model”. In Proc. of the ACL 2012 System Demonstrations, pages 37–42.

[2] Banchs R E, (2012), “Movie-DiC: A movie dialogue corpus for research and development”. In Proc. of the 50th Annual Meeting of the ACL.

[3] Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C. Lawrence Zitnick. (2015) Microsoft COCO captions: Data collection and evaluation server.

[4] D. Bahdanau, K. Cho, and Y. Bengio, (2015) “Neural machine translation by jointly learning to align and translate”. In Proceedings of the International Conference on Learning Representations (ICLR), San Diego, CA.

[5] M. Sundermeyer, R. Schluter, and H. Ney, (2010) “LSTM neural networks for language modeling”. In INTER SPEECH.

[6] J. Devlin, R. Zbib, Z. Huang, T. Lamar, R. Schwartz, and J. Makhoul, (2014) “Fast and robust neural network joint models for statistical machine translation”, In ACL.

[7] Lifeng Shang, Zhengdong Lu, and Hang Li, (2015) “Neural responding machine for short-text conversation” In ACL-IJCNLP, pages 1577–1586.

[8] Ilya Sutskever, Oriol Vinyals, and Quoc V Le, (2014) “Sequence to sequence learning with neural networks”, In Advances in neural information processing systems (NIPS), pages 3104–3112.

[9] K. Yao and G. Zweig, (2015) “Sequence-to-sequence neural net models for grapheme-to-phoneme conversion”, In INTERSPEECH.

[10] Y. Bengio, P. Simard, and P. Frasconi, (1994) “Learning long-term dependencies with gradient descent is difficult”. IEEE Transactions on Neural Network.

[11] A. Graves, (2013) “Generating sequences with recurrent neural networks”.

[12] S. Hochreiter and J. Schmidhuber, (1997) “Long short-term memory. Neural Computation”.

[13] Sutskever, I., Vinyals, O., and Le, Q. V, (2014) “Sequence to sequence learning with neural networks”. In NIPS.

[14] Ankur Bapna, Gokhan Tur, Dilek Hakkani-Tur, and Larry Heck, (2017) “Sequential dialogue context modeling for spoken language understanding”. In Proc. of SIGDIAL

[15] Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus, (2015) “End-to-end memory networks”,.arXiv preprint arXiv:1503.08895.

[16] K. Cho, B. van Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, (2014) “Learning phrase representations using run encoder-decoder for statistical machine translation”.

[17] Markus Freitag, Yaser Al-Onaizan (2017), “Beam Search Strategies for Neural Machine Translation”. In Workshop on Neural Machine Translation.

[18] V. Serban, A. Sordoni, Y. Bengio, A. Courville, and J. Pineau (2015), “Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Networks”.

[19] V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, (2016) “A hierarchical latent variable encoder-decoder model for generating dialogues”.

[20] Werner Deutsch and Thomas Pechmann. (1982), “Social interaction and the development of definite descriptions”, Cognition, 11:159–184.

[21] Alfred Kobs. (1990), “User modeling in dialog systems Potentials and hazards”. In AI & society.

[22] William Yang Wang, Ron Artstein, Anton Leuski, and David Traum, (2011), “Improving spoken dialogue understanding using phonetic mixture models”. In FLAIRS Conference.

[23] Roberto Pieraccini, David Suendermann, Krishna Dayanidhi, and Jackson Liscombe (2009), “Are we there yet? Research in commercial spoken dialog systems”. In Text, Speech, and Dialogue.

[24] Kaisheng Yao, Geoffrey Zweig, and Baolin Peng (2015), “Attention with the intention of a neural network conversation model”. CoRR, abs/1510.08565.

