International Journal of Advanced Science and Technology
Vol. 29, No. 03, (2020), pp. 4931- 4938

Novel Self Learning Agent for Chatbot Applications

Akhilesh Kumar Sharma[1], Venkatesh Gauri Shankar[2], Narendra Singh Yadav[3], Parvathi Saxena[4], Samrit Kumar Maity[5],
[1,2,3,4] *School of Computing and Information Technology, Manipal University Jaipur,* [5]*C-DAC Pune,*
[1]*akhileshkumar.sharma@jaipur.manipal.edu,*

Abstract

With over 5 billion users using chat applications on a daily basis. This has spiked an interest of large organizations to leverage the power and strength of chat based applications. With the market being more inclined to using chat applications over other modes of communication applications, there has been an increased demand for creating chatbots to provide support over, purchasing, selling etc kinds of tasks. Creating a chatbot restricted to a set of queries within limited context and answers is an easy task but it gets more challenging when you require the bot to keep a conversation going without any human intervening in the process. In this paper we explain the approach through which we train a neural network model to learn and answer to questions on its own while keeping a track of variable context. Using bAbI dataset from Facebook we were able to successfully train our model to do the same with 89% accuracy on an average over 20 tasks provide in the bAbI dataset.

Keywords: *Chatbots, ML (Machine Learning), RNN (Recurrent Neural Network), CNN (Convolutional Neural Network), LSTM (Long-Short Term Memory), GRU (Gated Recurrent Unit).*

Introduction

In a recent survey it was found that people around the globe prefer using WhatsApp over Twitter and Facebook. The data collected from this survey concluded that chat applications keep the user engaged in conversation since some manual exertion is required from the user’s end too. Many organizations have ported their support over chatbots allowing the user to get quick responses to their queries which in turn removed the monotonous waiting for being assigned to support team for getting any query resolved. But this did not completely remove the requirement of human involvement in the process. Somewhere during the process human assistance is inevitable because the chatbot wouldn’t truly be able to keep the conversation going and simulate the real time interaction when the challenge is with multiple conversation context come into the picture..

To eliminate the need of human intervention in AI applications is a major priority in their applications. Chatbots are applications that mimic a human conversation but in text exchanging form. But creating them is very challenging. In a traditional approach chatbots were designed with AIML[21] scripts, where the programmer would think of all possible inputs and map them to a particular output. But this the robustness of chatbot were limited to the fact that how many diverse scenarios are under considerations, how those scenarios would be handled. In short it was limited according to the thought process of programmers. There would still be some input scenarios that is left unattended causing the chatbot to fail.

In today’s world, conversations between various parties became very inconsistent with all the SMS language and different slangs used in it is hard to train a bot on all these

ISSN: 2005-4238 IJAST

inconsistencies. This would also be a long and time consuming task to do so, and there will always be a slim chance of something being left out causing the bot to fail. Using a neural network model to learn how to converse was the right way to go excluding some small conversational aspects like greetings could be hardcoded into the bot application. But there were shortcomings to these models because the bot is only as smart as what is learnt upon. This is when scientists moved to deep learning models.

There is ambiguity present in conversations, they are unpredictable and inharmonious. RNN models were designed to handle variable-length data (sentences) and they had internal memories allowing them to remember inputs that would help to predict the next output. The challenge faced here was feeding the model with tokens from the input set was not enough as there was no true understanding of what they meant. So researchers moved to word embedding's which were able to establish relations between those tokens. When RNNs were fed with word embedding the relations between words was understood correctly.

In this paper we have aimed to understand what scenarios of a conversation fails for chatbot to learn on. This paper has been divided into six sections as follows, (1) introducing the area of interest, (2) covers the survey of previous works done in the area of interest, in brief (3) explains the proposed model, (4) the results of the model (5) conclusion of our paper and lastly the (6) references.

Related Work

Vinyals[3] et. al proposed a model created with two Seq-2-Seq (Sequence-to-Sequence) model where one was called the encoder while the other called a decoder. The encoder was responsible for generating context (meaningful content) from the word embeddings fed to it, while the decoder would be fed with the output of the encoder to generate relevant output i.e. the response to the query in a Bot application.

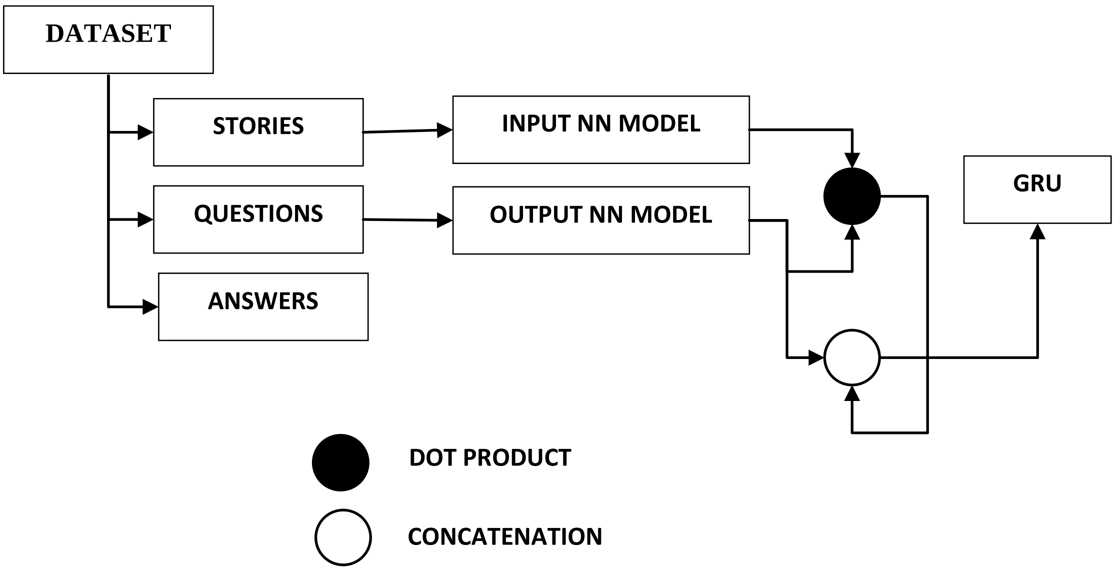

Figure 1. Flow Diagram

But in their model the responses were generic and different outputs would come for the same question. They suggested using SpeakerID as an added feature to the input which would in turn allow the bot to keep track of the flow of the conversation. Say any change in the context was faced then the SpeakerID would change in turn making the bot understand the change in context. Two major challenges identified by Liu[8] et. al were the mining of rich features from the dataset and complete utilization of label correlations.

ISSN: 2005-4238 IJAST

The model proposed by them was called Extreme Multi-Label CNN (XML-CNN) originally inspired by Kim-CNN[10]. Compared to a traditional CNN, this would swipe through the temporal aspect of the input allowing a larger coverage over a relevant subset of labels from different categories. Usage of word embeddings was emphasized in their work also. Using Max Pooling for keeping the context of the input intact was proposed by Kobayashi[17] et. al called the Dynamic Entity Relation model. A attention based CNN model was proposed by Yin[18] which failed when too many questions were asked. A CNN model[19] was designed to read the daily news and answer questions regarding it which scored an accuracy of 74%. Data Sparsity faced in the training process was addressed by You[11] et. al using BoW (Bag-of-Words) which allowed reducing the size of the label vectors to a smaller dimensional embedding in the target space. But this did not always work for large data sizes.

Deep Learning models were used for Multi-Label classifications giving very good results. Liu[12] et. al proposed a similar approach to extract features from raw text automatically and succeeded. Creation of Deep Learning models started with the traditional RNN[13] model. RNNs take two inputs, one being the present state and the other being the previous(recent past) state. These are very crucial inputs which tells the RNN what will happen next. Because of this learning form RNN stands out compared to other algorithms.

But RNNs fail sometimes to learn as some relevant information gets lost due to vanishing/exploding gradient problem. To solve this problem LSTMs(Long-Short Term Memory)[4,9] models were proposed. Keeping in mind that RNNs failed to keep relevant information in-memory, this was achieved using the tanh activation function. If the information being fed to this model was too large, then deciding what gets to stay in-memory and what gets discarded was hard to make[7]. This brought in the dawn of Dynamic Memory models which allowed only steady information to stay in-memory reducing the load on the LSTM model, allowing a smoother learning curve. The memory model consists of two memories called dynamic and episodic. Because LSTM works with two different activation functions i.e. tanh and sigmoid, helped in deciding what gets remembered and what gets forgotten as the outputs depended on either (0,1) or (1,1).

GRU was inspired from the LSTM model. Compared to the complex LSTM model with its various gates and activation functions, GRU combined the functionalities of these gates to only two gates namely the forget gate and update gate.

But GRU also couldn’t truly keep a conversation going. This was because if the context was changed it failed to keep up. To overcome this, SpeakerID [12] feature was introduced to help it track the context of a conversation. Using LSTM Tian[20] et. al proposed a ranking a technique to track the flow of a conversation. They were successful with some tasks and failed in inference, text summarization and event detection despite error analysis. The conclusion to their approach was splitting the task between different models.

Proposed Model

A serious problem faced by neural network models is that it fails to interpret relations between words fed to it. With Seq-2-seq model Liu [8] et. al were able to successfully achieve this with the usage of word embeddings. Vanishing/Exploding gradient problem was another problem faced by them. LSTM [3] was able to solve this. The remembering and forgetting mechanism was improved adding an external memory source proposed in the Dynamic Memory model [6], but only worked with smaller datasets.

ISSN: 2005-4238 IJAST

Table 1. Results

Challenge Type	LSTM	DM	Our Model
Single Supporting Fact	52.1	100	96
Two Supporting Facts	37	100	65.1
Three Supporting Facts	20.5	20	92.55
Two Arg Relations	62.9	71	93.9
Three Arg Relations	61.9	83	95.29
Yes No Questions	50.7	47	97.73
Counting	78.9	68	92.15
Lists Sets	77.2	77	83.19
Simple Negation	64	65	98.31
Indefinite Kknowledge	47.7	59	98.54
Basic Coreference	74.9	100	98.1
Conjunction	76.4	100	98.92
Compound Coreference	94.4	100	99.69
Time Reasoning	34.8	99	88.39
Basic Deduction	32.4	74	77.94
Basic Induction	50.6	27	59.4
Positional Reasoning	49.1	54	81.95
Size Reasoning	90.8	57	99.62
Path Finding	9.8	0	20.66
Agents Motivation	90.7	100	98.95

Our proposed model consists of two sequential models and a classic GRU model. The two sequential models are responsible for handling the input. Since we are using the bAbI dataset, it consists of a story, questions of that story and answers to that question. Say our chatbot has three functional modules namely (1) understanding, (2) input and (3) output. Understanding would be represented by the story in the dataset as it tells us what it learns. The input would be the question queried by the user represented by the questions of that story and the output would be the reply made by the bot represented by the answers to the questions. bAbI dataset consists of 20 tasks each representing a different area of a conversation, like deductive, inductive, set etc.

The two sequential models will be used to handle the stories and questions each. The dot product of the outputs of each model and then fed to a softmax activation function so that the probabilities of choosing the next word would be achieved. This output is then added to the output of the questions handling sequential model for mapping sake. The response generated is then fed to GRU as an input for training.

Result

We compared our model results with other models like LSTM[1] and Dynamic Memory model(DM)[2].

ISSN: 2005-4238 IJAST

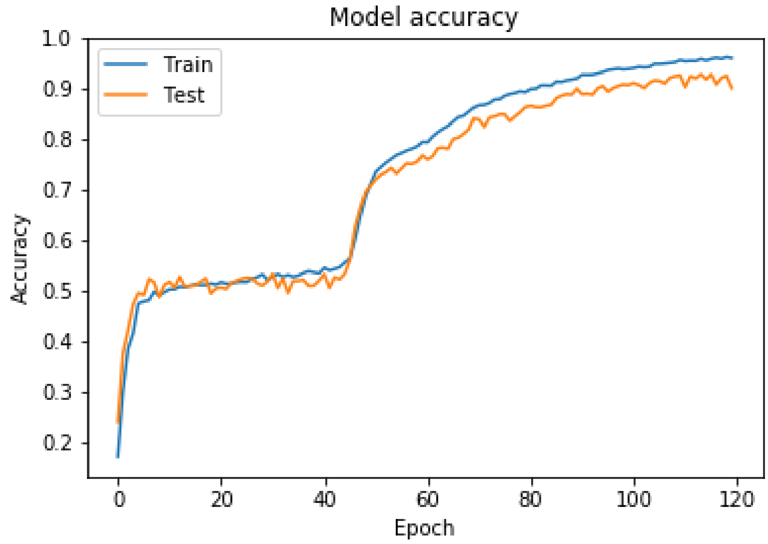

Figure 1. Model Accuracy of single supporting fact for 120 epochs

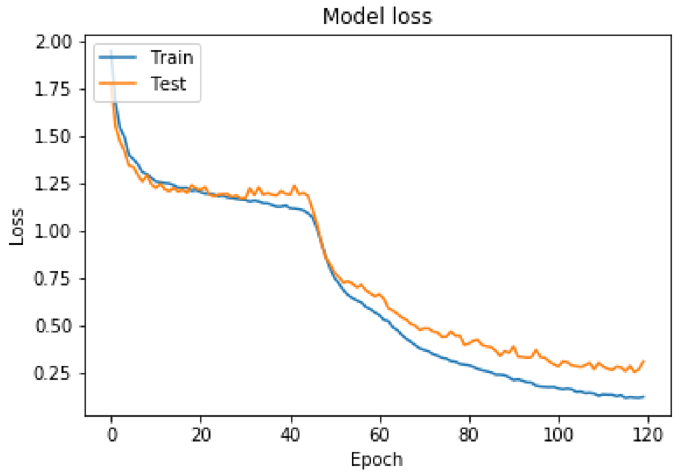

Figure 2. Model loss for single supporting fact for 120 epochs

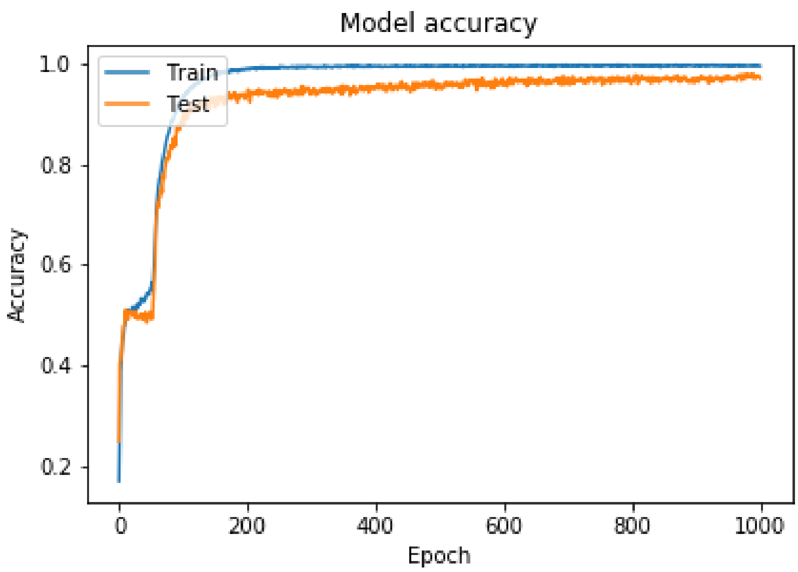

Figure 4. Model Accuracy of single supporting fact for 1000 epoch

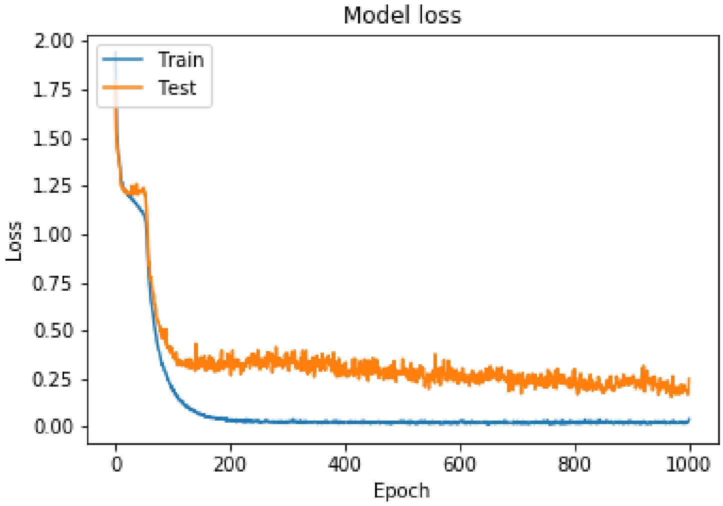

Figure 5. Model loss of single supporting fact for 1000 epoch

The results of these models are taken from the previous work [1,2]. For training, the number of epochs were set to 120 and the batch size was 32 (default). Looking at the first task i.e. Single Supporting Fact, our model scored 96% accuracy with 98.7% for training and 90.8% for testing. We then increased the number of epochs to 1000 to look for overfitting condition which does not exist.
With Path Finding task, we looked into the dataset and found that the directions were encoded into specific characters like 's' for south, 'n' for north etc. We probably need some mechanism that will help our model to understand these encodings. With the Basic Induction task our model fails to understand the hierarchical nature of the story and we would require some tree-based model for these kinds of input.

Conclusion

Usage of neural network models for creating chatbot applications has improved the learning process largely. Our model was able to score about 86% accuracy on an average across the 20 tasks. For the tasks that it failed on we could try using Dynamic Co-attention network model [14,15] for improving the learning curve for those tasks.

References

[1] Weston, J., Bordes, A., Chopra, S., Rush, A.M., van Merriënboer, B., Joulin, A. and Mikolov, T., 2015. Towards ai-complete question answering: A set of prerequisite toy tasks. arXiv preprint arXiv:1502.05698.
[2] Liu, F., Cohn, T. and Baldwin, T., 2017, December. Improving End-to-End Memory Networks with Unified Weight Tying. In Proceedings of the Australasian Language Technology Association Workshop 2017 (pp. 16-24).
[3] Vinyals, O. and Le, Q., 2015. A neural conversational model. arXiv preprint arXiv:1506.05869.
[4] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. Neural computation, 9(8), pp.1735-1780.
[5] Shen, G., Yang, Y. and Deng, Z.H., 2017, September. Inter-weighted alignment network for sentence pair modeling. In Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (pp. 1179-1189).
[6] Raghuvanshi, A. and Chase, P., 2016. Dynamic memory networks for question answering.
[7] Bhagwat, Vyas Ajay., 2018. Deep Learning for Chatbots.
[8] Liu, J., Chang, W.C., Wu, Y. and Yang, Y., 2017, August. Deep learning for extreme multi-label text classification. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 115-124). ACM.
[9] Su, M.H., Wu, C.H., Huang, K.Y., Hong, Q.B. and Wang, H.M., 2017, December. A chatbot using LSTM-based multi-layer embedding for elderly care. In 2017 International Conference on Orange Technologies (ICOT) (pp. 70-74). IEEE..

ISSN: 2005-4238 IJAST

[10] Kim, Y., 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[11] You, R., Dai, S., Zhang, Z., Mamitsuka, H. and Zhu, S., 2018. Attentionxml: Extreme multi-label text classification with multi-label attention based recurrent neural networks. arXiv preprint arXiv:1811.01727.

[12] Liu, J., Chang, W.C., Wu, Y. and Yang, Y., 2017, August. Deep learning for extreme multi-label text classification. In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 115-124). ACM.

[13] Gers, F.A., Schmidhuber, J. and Cummins, F., 1999. Learning to forget: Continual prediction with LSTM.

[14] Kadlec, R., Schmid, M., Bajgar, O. and Kleindienst, J., 2016. Text understanding with the attention sum reader network. arXiv preprint arXiv:1603.01547.

[15] Weston, J., Chopra, S. and Bordes, A., 2014. Memory networks. arXiv preprint arXiv:1410.3916.

[16] Dhingra, B., Liu, H., Yang, Z., Cohen, W.W. and Salakhutdinov, R., 2016. Gated-attention readers for text comprehension. arXiv preprint arXiv:1606.01549.

[17] Kobayashi, S., Tian, R., Okazaki, N. and Inui, K., 2016, June. Dynamic entity representation with max-pooling improves machine reading. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (pp. 850-855).

[18] Yin, W., Ebert, S. and Schütze, H., 2016. Attention-based convolutional neural network for machine comprehension. arXiv preprint arXiv:1602.04341.

[19] Chen, D., Bolton, J. and Manning, C.D., 2016. A thorough examination of the cnn/daily mail reading comprehension task. arXiv preprint arXiv:1606.02858.

[20] Tian, T. and Li, Y., 2016. Machine Comprehension Based on Learning to Rank. arXiv preprint arXiv:1605.03284

Authors

Dr. Akhilesh Kumar Sharma, Currently working in Manipal University Jaipur, India as Associate prof., Ph.D. from JK-SPS University, Udaipur in CSE, M.E. in CSE from DAVV University (Indore), has done B.E. in CSE from RGTU. He has won the best research proposal award by ACM and CSI Udaipur chapter. His area of interests include Big data analytics, deep learning, classification and clustering methods, Data mining and machine learning, cognitive intelligence etc. He has many technical research papers in international and national journals conferences. He has visited several places for presenting research papers, chairing sessions and delivering Expert lectures and keynotes in IITs, NITs, Vietnam, Thailand, Malaysia and China etc. He is a Senior Member of IEEE, ACM, CSI, (IUCEE), and Member compliance committee of MIR Labs Washington, USA. He has organized conferences, faculty development programs, workshops and seminars with GOI funding. Due to his meritorious contributions in the field of education, he has been awarded Best paper Award, NTU Singapore invited as visiting research scientist etc. He has received Best teacher award for his contributions in academics. Currently he is joint secretory in ACM professional chapter Jaipur and ACM student chapter faculty coordinator & won best student chapter award for MUJ under his able leadership. He also got patent published and has setup cognitive intelligence research lab in Manipal university Jaipur, Jaipur, Rajasthan.

Mr. Venkatesh Gauri Shankar is working as Assistant Professor in Department of Information Technology, School of Computing and IT, Manipal University Jaipur, India. He has completed his MCA from Uttar Pradesh Technical University and M Tech. in Computer Science

ISSN: 2005-4238 IJAST

and Engineering from Central University of Rajasthan. He has published more than 30+ papers in reputed journals and conferences indexed in Scopus and Web of science. He has total of 11+ years teaching experience in academics. He got best paper, best reviewer and best innovator award by many professional bodies like Springer, CRC etc.. He has also reviewer and board member of many reputed journals of Elsevier and IEEE. He has also served as session chair and advisory board in many national and international conferences. He is the member of many professional bodies like IEEE, Springer, CSI, ACM etc. He has also published patent for diabetic treatment using Artificial Intelligent. His one project named “HandyDP: Hand Gesture Recognition for Deaf and Dumb” is approved and funded by DST Rajasthan. He has also visited and having research collaboration with Malaysia, Singapore, Thailand, Vietnam and Italy. His research and specialized areas are centric to Machine Learning, Deep Learning and Data Science. Currently working in the area of many biomedical and neurodegenerative applications of Deep Learning and Machine Learning.

Dr. Narendra Singh Yadav received his M.Tech. degree in Computer Science from Birla Institute of Technology, Ranchi and Ph.D degree from Malaviya National Institute of Technology, Jaipur. He is currently an Associate Professor (Senior Scale) of Information Technology with Manipal University Jaipur, Jaipur. His research interests include Computer Networks, Network Security, Digital Forensics, Cyber Security, Malware and Reverse Engineering. He is an active member of various professional bodies and has published over 50 papers in International/national Journals/conferences. He is a certified ethical hacker (CEH), certified hacking forensics investigator (CHFI), certified ec-council instructor (CEI) and ISO 27001 LA.