Making Bots Smarter Using Memory Networks

Deep Narain Singh

M.S Data Science : GalvanizeU, University Of New Haven
44 Tehama St., San Francisco, CA
singh.deep@gmail.com

August 12, 2016

Abstract

We want to build Dialogue agents or Conversational agents which is intelligent enough to have conversation with human.We want the bot to attain intelligence level which can handle long and open ended conversation and learn from the conversation and can understand the long term context. In order to build intelligent end to end dialogue agent and to move towards our goal we experimented with new kind of model called as memory network which was introduced by Facebook Research Team (Jason Weston and team).Most of the neural network model has short term memory and in order to understand long term conversation or dialogue and provide response to that the model needs to have a long term memory. To measure progress towards our goal we will experiment with the RNN/LSTM neural network and compare its performance with the memory component added to it. This experiment is conducted on bAbi dataset which is provided by Facebook Research Team.

1 Introduction

One of the long term goal of NLP and AI is to build a conversational agent which is intelligent enough to converse with human and can understand the conversation like human do and provide appropriate response.In order to build such a dialog agent first let's think what and ideal dialogue agent could be ?
An ideal agent is one which can use its pre-existing knowledge to perform complex tasks , it should be able to understand context and motives in long term conversation and respond appropriately it should also keep on learning as it converse and further increase its knowledge and capability.

One of the challenges in building these agents is understanding the context in long term conversation and provide answer based on that long conversation.Such type of system needs to remember different states and understand the motive and language understanding in underlying conversation .
Other challenge is how do we evaluate the performance of such a dialog agent in natural conversation and developing a mechanism to further improve its performance.

There is lot of research work going on to build such a intelligent agent and to solve these problems. One of the interesting approach is to imagine these system as Question,Answering system, the task of question answering fits the evaluation criteria for dialogue agent because agent response to question can be evaluated to the expected answer.

Facebook Research has done some interesting work in this domain and they have released bAbi data set which can be evaluated on 20 different type of task like inference , negation , counting etc. In order to solve the problem of having memory in long term conversation we will experimenting a new type of model called as Memory Network which is also released by Facebook AI Research Team. These models combine large memory with learning component that can read and write to it.

2 Background

In any dialogue conversation it is important to remember the previous dialogues which can be short term conversation or long term conversation and then respond.We need a long term memory to remember the dialogue or story and then answer the questions.

It has been seen that Deep Neural Networks use RNN and LSTM as memory units but they are not efficient enough to understand long term conversation and make inference as they miss the structured memory component.

Memory Networks (Weston et al) is a model which combines large memory with learning component , the memory acts as a knowledge base which can recollect the facts from the past and make a inference. Memory networks combine inputs with attention on memories to provide the reasoned outputs.

3 Theory Of Solution

3.1 Problem Definition

Our goal is to build an agent which can answer questions asked by humans in natural conversation.In this task there will be a sequence of sentences in form of story or facts which will be used for building knowledge base . There will be questions which will be asked to agent based on the story . Agent is expected to understand the questions and give the best possible answer . Below is a example of sample story , a question and an answer.

John was in the bedroom.
Bob was in the office.
John went to the kitchen.
Bob travelled back home.
Where is John? A: **kitchen**

3.2 Dataset

We will be using **bAbi** dataset provided by facebook research team to evaluate the performance of memory networks. This dataset contains 20 different type of task of varying difficulty level and test different memory and inference skill of the dialog agent. This dataset consist of set of sort stories.Each story is a sequence of statements about an evolving situation. The model should read a single story and answer one or more questions about it. Within a benchmark, the stories test the same skill. Across the different benchmarks, the skills get more difficult. Below is the table representing 20 different types of task.

Table 1: Sample statements and questions from tasks 1 to 10.

Task 1: Single Supporting Fact Mary went to the bathroom. John moved to the hallway. Mary travelled to the office. Where is Mary? A:office	**Task 2: Two Supporting Facts** John is in the playground. John picked up the football. Bob went to the kitchen. Where is the football? A:playground
Task 3: Three Supporting Facts John picked up the apple. John went to the office. John went to the kitchen. John dropped the apple. Where was the apple before the kitchen? A:office	**Task 4: Two Argument Relations** The office is north of the bedroom. The bedroom is north of the bathroom. The kitchen is west of the garden. What is north of the bedroom? A: office What is the bedroom north of? A: bathroom
Task 5: Three Argument Relations Mary gave the cake to Fred. Fred gave the cake to Bill. Jeff was given the milk by Bill. Who gave the cake to Fred? A: Mary Who did Fred give the cake to? A: Bill	**Task 6: Yes/No Questions** John moved to the playground. Daniel went to the bathroom. John went back to the hallway. Is John in the playground? A:no Is Daniel in the bathroom? A:yes
Task 7: Counting Daniel picked up the football. Daniel dropped the football. Daniel got the milk. Daniel took the apple. How many objects is Daniel holding? A: two	**Task 8: Lists/Sets** Daniel picks up the football. Daniel drops the newspaper. Daniel picks up the milk. John took the apple. What is Daniel holding? milk, football
Task 9: Simple Negation Sandra travelled to the office. Fred is no longer in the office. Is Fred in the office? A:no Is Sandra in the office? A:yes	**Task 10: Indefinite Knowledge** John is either in the classroom or the playground. Sandra is in the garden. Is John in the classroom? A:maybe Is John in the office? A:no

Table 2: Sample statements and questions from tasks 11 to 20.

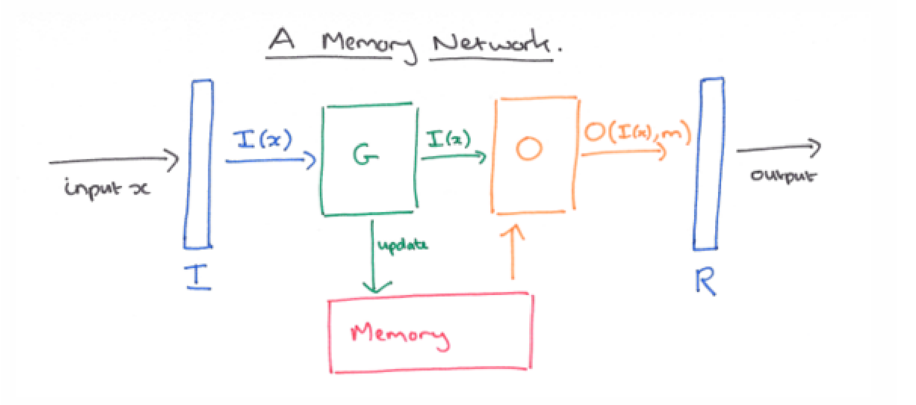
Task 11: Basic Coreference Daniel was in the kitchen. Then he went to the studio. Sandra was in the office. Where is Daniel? A:studio	**Task 12: Conjunction** Mary and Jeff went to the kitchen. Then Jeff went to the park. Where is Mary? A: kitchen Where is Jeff? A: park
Task 13: Compound Coreference Daniel and Sandra journeyed to the office. Then they went to the garden. Sandra and John travelled to the kitchen. After that they moved to the hallway. Where is Daniel? A: garden	**Task 14: Time Reasoning** In the afternoon Julie went to the park. Yesterday Julie was at school. Julie went to the cinema this evening. Where did Julie go after the park? A:cinema Where was Julie before the park? A:school
Task 15: Basic Deduction Sheep are afraid of wolves. Cats are afraid of dogs. Mice are afraid of cats. Gertrude is a sheep. What is Gertrude afraid of? A:wolves	**Task 16: Basic Induction** Lily is a swan. Lily is white. Bernhard is green. Greg is a swan. What color is Greg? A:white
Task 17: Positional Reasoning The triangle is to the right of the blue square. The red square is on top of the blue square. The red sphere is to the right of the blue square. Is the red sphere to the right of the blue square? A:yes Is the red square to the left of the triangle? A:yes	**Task 18: Size Reasoning** The football fits in the suitcase. The suitcase fits in the cupboard. The box is smaller than the football. Will the box fit in the suitcase? A:yes Will the cupboard fit in the box? A:no
Task 19: Path Finding The kitchen is north of the hallway. The bathroom is west of the bedroom. The den is east of the hallway. The office is south of the bedroom. How do you go from den to kitchen? A: west, north How do you go from office to bathroom? A: north, west	**Task 20: Agent's Motivations** John is hungry. John goes to the kitchen. John grabbed the apple there. Daniel is hungry. Where does Daniel go? A:kitchen Why did John go to the kitchen? A:hungry

3.3 Memory Networks

A memory network combines learning strategies from the machine learning literature with a memory component that can be read and written to. The model is trained to learn how to operate effectively with the memory component. The high-level view of a memory network is as follows:

- There is a memory, m, an indexed array of objects (e.g. vectors or arrays of strings).
- An input feature map I, which converts the incoming input to the internal feature representation
- A generalization component G which updates old memories given the new input. "We call this generalization as there is an opportunity for the network to compress and generalize its memories at this stage for some intended future use."
- An output feature map O, which produces a new output in the feature representation space given the new input and the current memory state.
- A response component R which converts the output into the response format desired – for example, a textual response or an action.

I,G,O and R can all potentially be learned components and make use of any ideas from the existing machine learning literature. They are connected as follows:

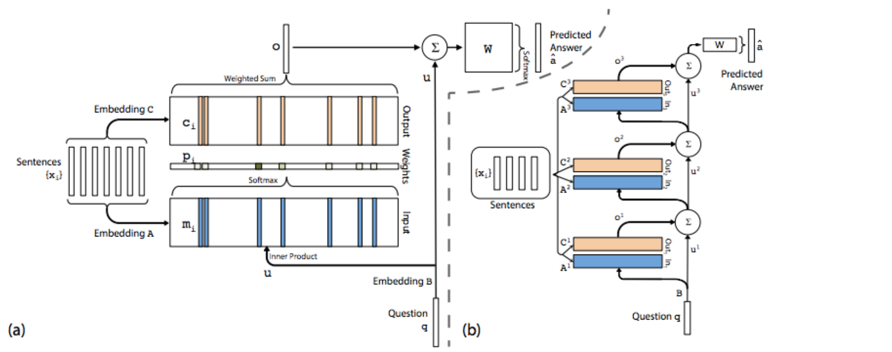


When the components I,G,O and R are neural networks, the resulting system is a Memory Neural Network (MemNN). We build MemNN for QA (question answering) problems and compare it to RNNs (Recurrent Neural Network) and LSTMs (Long Short Term Memory RNNs) and find that it gives superior performance. Below is the proposed architecture and its explanation

1. Input Memory Representation: Let the input sequence be $x_i = x_1, x_2, ...x_i$ Each word of the sentence is embedded using a matrix A. For each sentence a corresponding memory m_i by summing up the embeddings of all the words in a sentence.
   $m_i = \sum_j Ax_{ij}$
   The question is embedded using another matrix B and we calculate the match between the question and each memory as
   $p_i = softmax(u^T m_i) = softmax(q^T B^T \sum_j Ax_{ij})$

2. Output Memory Representation: Each memory vector has a corresponding output vector c_i computed using another matrix C. $c_i = \sum_j Cx_{ij}$
   The output vector from the memory o is then summation of c_i weighted by the probability vector from the input.
   $o = \sum_i p_i c_i$

3. Generating Final Prediction: In the single layer case, the sum of the output vector o and the input embedding u is then passed through a final weight matrix W (of size V $\times$ d) and a softmax to produce the predicted label:
   $\hat{a} = Softmax(W(o + u))$



4 Implementation

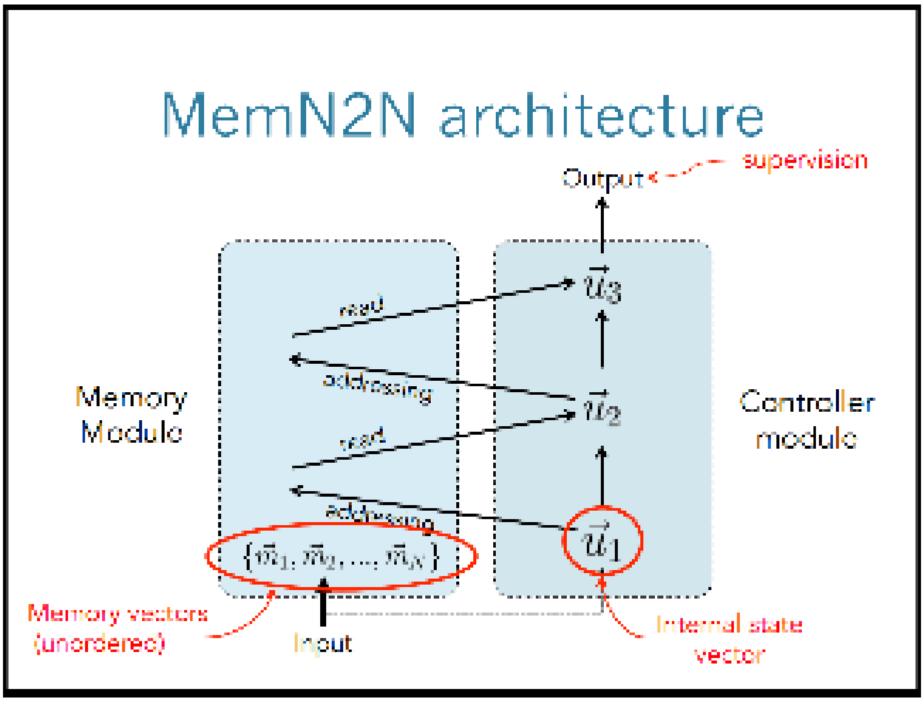
- Tools and Library: We use tensorflow , python 2.7, numpy etc to build this prototype.

- Data Processing: bAbi dataset from facebook is used. Each task has a story component and query component.Data is preprocessed to convert sequence of words and query into vector representation. The fixed vector representation is feeded to the neural network and hope that it answers correctly.

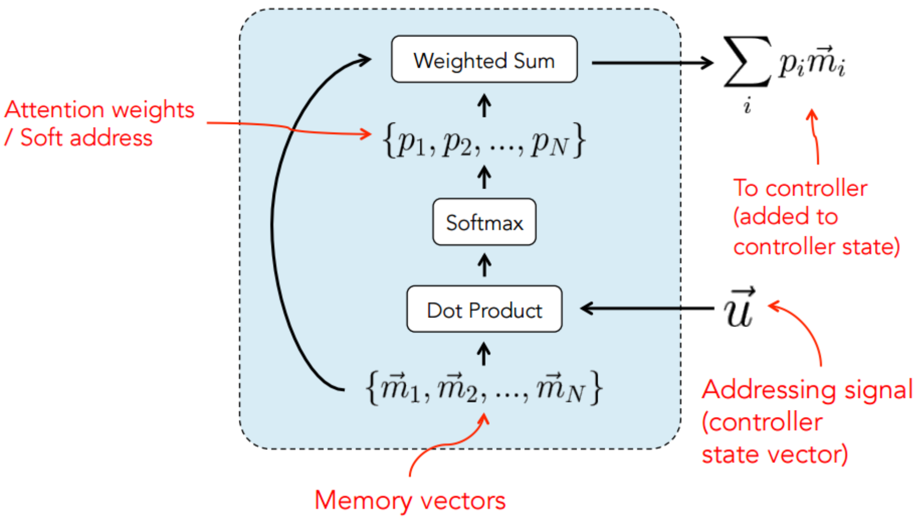
- Training Steps: Below are the training steps for neural network model with external memory.

  1. Reads from a memory with soft attention.
  2. Perform multiple lookups or hops on memory.
  3. End to End training with backpropogation.

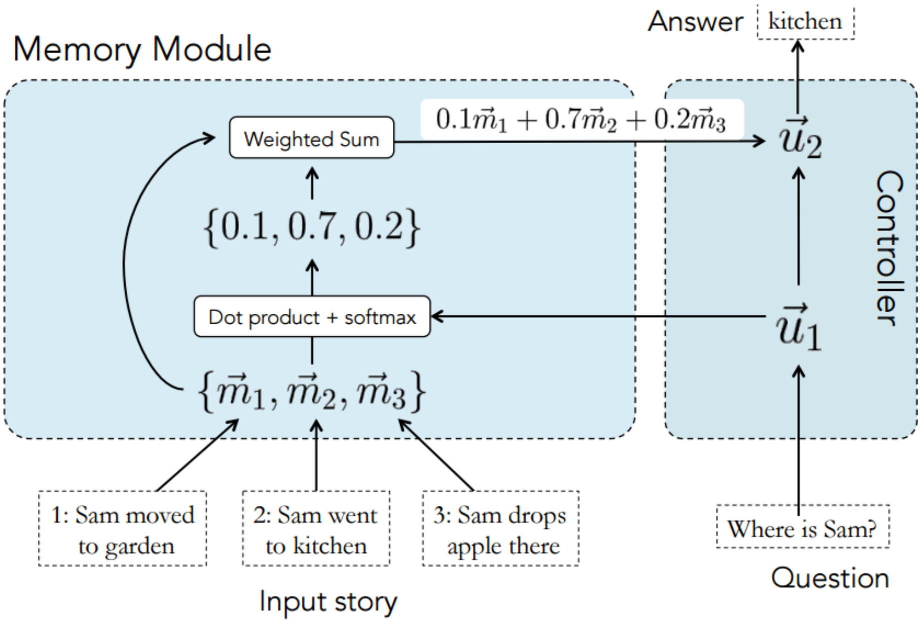
- Building Network : There are 2 modules in MemN2N architecture the memory module and the controller module. First the input story vector is feeded into memory vector and Controller maintains the internal state, it can come from input or can be fixed vector. First you will do a look up from state to memory vector then both can be added to get the next state vector.This process can be repeated several times.Finally decoder is used to get the final output.To train this whole model we need supervision only on the output.



- Memory Module : There are memory vectors in memory and addressing signal comes from the controller or the state vector. We take the dot product between those two vectors.It will give a similarity score through softmax and give probability distribution over memory location.It can also be considered as attention weights.Then compute the weighted sum of memory vectors. The output will go the controller and added to the current controller state.



- Experiment on Question Answering: Suppose we have three sentences as input , we will feed each sentence into a memory vector.The query or question vector will give the controller state.We apply a dot product and apply softmax to get the attention weights.After that we will do a weighted sum and add vector to the controller state and decode that to get the answer.



5 Result

We evaluate the benchmark on the number of predictions accurately answered.We see the accuracy rate for each of the task on the test dataset. When Result is compared with the LSTM baseline published by facebook research team it is seen that Memory Network outperformed the LSTM based model. If the accuracy score is more than 94 percent the task is considered to be passed.It was seen that in LSTM based model number of failed task are 19 while in simple MemN2N model failed task were 11. Below is the result.

Task	FAIR LSTM Baseline Score	MemN2N Score
Single Supporting Fact	50	96
Two Supporting Facts	20	87
Three Supporting Facts	20	60
Two Arg. Relations	61	97
Three Arg. Relations	70	87
Yes/No Questions	48	92
Counting	49	83
Lists/Sets	45	90
Simple Negation	64	87
Indefinite Knowledge	44	85
Basic Coreference	72	95
Conjunction	74	96
Compound Coreference	94	96
Time Reasoning	27	97
Basic Deduction	21	96
Basic Induction	23	95
Positional Reasoning	51	49
Size Reasoning	52	89
Path Finding	8	7
Agent's Motivations	91	95
Failed Task	19	11

We can also visualize the attention during the memory of hops ,no of hops is a hyperparameter which we can optimized.

Story (1: 1 supporting fact)	Support	Hop 1	Hop 2	Hop 3
Daniel went to the bathroom.		0.00	0.00	0.03
Mary travelled to the hallway.		0.00	0.00	0.00
John went to the bedroom.		0.37	0.02	0.00
John travelled to the bathroom.	yes	0.60	0.98	0.96
Mary went to the office.		0.01	0.00	0.00
Where is John? Answer: bathroom Prediction: bathroom				

Story (2: 2 supporting facts)	Support	Hop 1	Hop 2	Hop 3
John dropped the milk.		0.06	0.00	0.00
John took the milk there.	yes	0.88	1.00	0.00
Sandra went back to the bathroom.		0.00	0.00	0.00
John moved to the hallway.	yes	0.00	0.00	1.00
Mary went back to the bedroom.		0.00	0.00	0.00
Where is the milk? Answer: hallway Prediction: hallway				

Story (16: basic induction)	Support	Hop 1	Hop 2	Hop 3
Brian is a frog.	yes	0.00	0.98	0.00
Lily is gray.		0.07	0.00	0.00
Brian is yellow.	yes	0.07	0.00	1.00
Julius is green.		0.06	0.00	0.00
Greg is a frog.	yes	0.76	0.02	0.00
What color is Greg? Answer: yellow Prediction: yellow				

Story (18: size reasoning)	Support	Hop 1	Hop 2	Hop 3
The suitcase is bigger than the chest.	yes	0.00	0.88	0.00
The box is bigger than the chocolate.		0.04	0.05	0.10
The chest is bigger than the chocolate.	yes	0.17	0.07	0.90
The chest fits inside the container.		0.00	0.00	0.00
The chest fits inside the box.		0.00	0.00	0.00
Does the suitcase fit in the chocolate? Answer: no Prediction: no				

6 Conclusion

We see that simple memory augmented neural networks on QA task outperform the LSTM based model. With more hyperparameter optimization we can further increase the performance.
We are still far away in process of building intelligent end to end dialogue agent but this approach fills the problem of memory and needs further enhancement. This experiments sets the direction of approach and validates our hypothesis that at present this is the right approach to improve the memory problem in bots.

6.1 Future Work

This experiment works well in structured data but in real world scenario most of the times the data will be unstructured. There is a recent paper released by facebook known as key-value memory network which works on unstructured data. We need to evaluate its performance and see if the model which has knowledge base in form of unstructured data then how well it performs on the queries. This work is in progress and needs to be evaluated.

7 Acknowledgement

I would like to thank my mentors Brian Spiering , Motoki Wu , David Bernat for their exceptional guidance and support.
Special thanks to Brian Spiering , Jared Thompson and Mike Bowles to build the foundation in NLP , Machine Learning and deep learning which was extremely useful during this project.
I would also like to thank Nir Kaldero ,the entire faculty and staff at Galvanize for their support.

8 References

[1] Weston, Jason , et al. ”Towards AI-complete question answering: A Set of prerequisite toy tasks.” arXiv preprint:arxiv:1502.05698v10

[2] Weston, Jason Sumit Chopra, and Antoine Bordes.”Memory Networks:”arXiv preprint:arxiv:1410.3916v11