International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

Dynamic Memory Inference Network for Natural Language Inference (2017)

Rama Krishna Raju

Abstract: ***In this paper, the NLI task is viewed as a question - and - answer problem, which leads to proposing the application of DMNs to enhance performance. NLI classifies the relationship between two sentences as entailment, contradiction, or neutral. It is necessary for many NLP applications, including question - answering, text summarization, and information retrieval. The most significant contribution of the paper is to demonstrate that DMNs are effective for episodic memory and experiment with DMNs outside of their original domain. This feature allows the model to incrementally update and reawaken memory, which provides a more nuanced view of the interactions of the sentences and hence also improves the accuracy of its inferential process. Furthermore, the paper analyses the aspects of the employed attention mechanisms in the structure of DMNs that enhance the capability of the model to pay attention to the essential words and phrases for successful task completion. The study highlights the impact of integrating the episodic memory updates with the attention mechanisms due to extensive experimentation, showing the benefits of such an NLI improvement and the potential for enhancing other NLP tasks.***

Keywords: Natural Language Inference, Dynamic Memory Networks, Attention

1. Introduction

Natural Language Inference, or NLI, can be defined as a significant task within the NLP field, which aims at identifying and recognizing the logical consequences implied by the provided pair of sentences. Typically, these relationships are categorized into affirming, negation, and related. This may mean if one of the two is true, the other cannot be false; this is commonly referred to as entailment. This refers to the situation where one sentence correlates with the paradox of the other or the truth of one does not impact or is unsure of the other. Because of its importance in NLP, NLI also represents a crucial control task for machine learning's semantic understanding abilities. The use of NLI is diverse and ranges in areas such as question answering automatic summarization, and information retrieval. Therefore, future progress in NLI can highly benefit general NLP systems.

At the beginning of NLI studies, much of the early work focused on shallow forms of representation. These methods worked on keywords and phrases and depended on the logic rule to establish correspondences between two sentences. A second attempt at defining logical patterns based on such patterns of inclusion was made more formal and syntactic in approach, called natural logic. However, while these approaches constituted state of the art for their time, there existed significant limitations in handling human language where the content may have multiple embedded interpretations or instances of irony or sarcasm.

Introducing deep learning has created higher models and increased workflow in Natural Language Understanding. For instance, LSTM gave the models the power to capture long - range dependencies within the sentence, which perfectly fits tasks like NLI, in which sequence information plays a crucial role. Another type of network was also helpful, Convolutional Neural Networks (CNNs), which were initially used for image processing but also entered text analysis by capturing local patterns. TREE - structured neural networks were the second analysis method as they also retain the hierarchical structure in the sentences, which is beneficial for syntax - related applications. Most importantly, attention - based models have been gaining grounds where models can attend to only constituent parts of the figure, which has boosted the performance of NLP tasks such as NLI.

Numerous existing deep learning architectures initially needed to incorporate memory mechanisms, which constitutes an essential problem for tasks that involve reasoning and iterative computations. Memory mechanisms enable models to store and update information from step to step, replicating how humans approach these problems. This capability becomes crucial in NLI since hearing the following sentence may force the reader to go back and reinterpret earlier sentences and see how they relate. It may be less beneficial for models to understand the whole relations between sentences, and investment in an effective memory mechanism may need to be revised.

Given these challenges, Dynamic Memory Networks (DMNs) can improve NLI models. DMNs can update their memory states iteratively through the interaction between sentences, so they are highly optimized for step - by - step reasoning tasks (Michael et al., 2017). The DMNs allow for the information that was previously received to be reevaluated and reconstructed, enabling the improved production of conclusions when compared to stationary models. This type of iterative process is most helpful in contexts such as NLI because how sentences relate to one another is sometimes complex and may take many cycles of passes over data to discern.

In this paper, we propose using DMNs to solve NLI, considering it a question - answering problem, where the premise would act as the context and the hypothesis as the question. We show this by showing how episodic memory updates with the DMN lead to achieving outstanding performance in inference from the two sentences as compared to the model that does not update its focus and is bound to make incremental updates to the understanding of the two sentences. The experiments we present in this paper demonstrate that this approach improves the model performance on standard NLI benchmarks and sheds light on when and how to use memory and attention mechanisms for solving other NLP tasks efficiently.

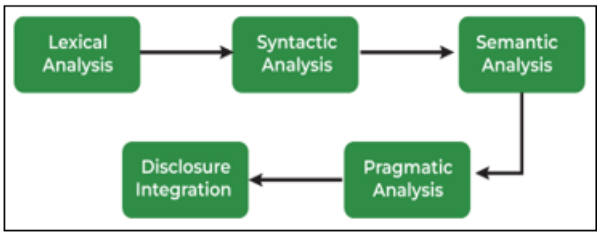


Figure 1: Phases of Natural Language Processing (NLP)

2. Related Work

Deep Learning for SNLI

The Stanford Natural Language Inference (SNLI) dataset has been essential in developing NLI, serving as a benchmarked corpus of labeled sentence pairs for model training and evaluation. Bowman, Angeli, et al. (2015) probably remain the only authors who employed deep learning techniques in this dataset. Training unlexicalized and lexicalized features supervised classifiers were their work, and the unlexicalized features yielded 50.4% accuracy compared to 78.2% of accurate lexicalized features. This approach was the first to highlight that feature selection is a critical issue for the NLI tasks. Besides these classifiers, the eight researchers proposed a new sentence encoding model using a 100 - dimensional LSTM encoder. This model trained each sentence individually and then joined the resultant embeddings from each of these training sessions, which was fed into a three - layered MLP for classification, and it performed with 77.6% accuracy. When The hidden size was made of 300 dimensions, accuracy was recorded at 80.6%, proving the effectiveness of dimensionality on the improved quality of feature vectors on the overall NLI task.

Further work extended from here, seeking to vibrate even more complex approaches to sentence encoding. For example, Mou et al. (2016) developed a tree - structured composition model that integrated local feature extraction with much simpler pooling methods. This was particularly useful in capturing some hierarchical structures of sentences that are useful in analyzing syntactic relations. Through targeted construction of sentence features, the proposed method further improved the understanding of how structural representations can be used to strengthen NLI. These early models provided the foundation for adding higher - level neural structures, like attention and memory, that can help achieve the best performance of NLI (Bowman, 2016).

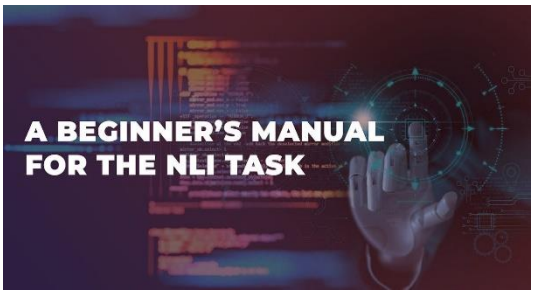


Figure 2: Natural Language Inferencing (NLI) Task

Neural Attention for SNLI

Neural attention has revolutionized natural language processing, which allows models to boost only the essential parts of the input data. Using this concept, Bahdanau, Cho, and Bengio (2015) demonstrated how attention mechanisms could be employed to decode and translate sequences of information in a much better way than was possible with traditional architectures. Constructing upon this approach, Rocktäschel et al. (2015) do something similar to the NLI task in using neural attention. Instead, they suggested building an encoding model that applied attention over a linear LSTM encoder to generate the hypothesis representation conditioned on the premise. More specifically, their method entailed feeding the hypothesis through a linear RNN that would attend to the states that the premise LSTM produced. The last attention state and the hypothesis hidden state were then passed into a classifier, considerably improving performance.

Building on that concept, Wang and Jiang (2015) have made some modifications by substituting the disused RNN with LSTM in the attention mechanism. This modification was beneficial in approximating long sequences of the premise and hypothesis better, thus correcting some of the drawbacks observed in previous models. Cheng, Dong, and Lapata (2016) elaborated on these ideas through intra - sentence attention, allowing the model to pay attention to the hidden states of tokens already processed in the current sentence. Hence, this self - attention mechanism on the within - sentence level extended a depth - expressed initialization of the hypothesis and the premise, providing better inference validity.

Such improvements to attention mechanisms have been extended to a wide array of other NLP tasks apart from NLI, such as image caption generation (Xu et al., 2015; Vinyals et al., 2015), machine translation, and question answering. Throughout these tasks, we observe that models with attention have shown generally higher performance than native LSTM and RNN models, proving the effectiveness of attention in effectively capturing sequence relations and dependencies. Specifically, the attention mechanisms have shown to be very useful in NLI, where identifying the complex semantics between the sentences is paramount.

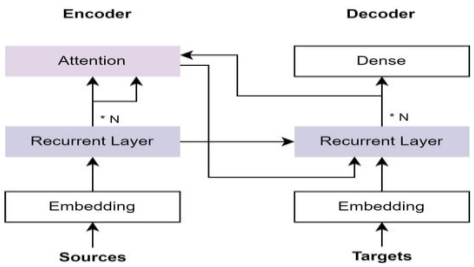


Figure 3: S2S encoder–decoder with attention architecture.

End - to - End Memory Networks

Memory networks have recently been shown to be highly effective models for tasks that entail series operations, notably NLI. Sainbayar et al. (2015) proposed an end - to - end memory - based model capable of performing multiple memory hops that can be learned using basic gradient descent techniques. This model performs better than baseline methods based on experiments conducted on QnA benchmark and language modeling. The architecture of the memory network enabled the system to focus on the most critical sentences at each time step to perform a computation to come up with the correct predictions.

The memory network proposed by Sainbayar et al. is structurally very similar in concept to the seq2seq attention model proposed by Bahdanau et al. (2015) but works at the sentence level instead of the token level. Besides, it uses a much more straightforward scoring function, which is helpful for tasks that require reasoning at the level of a single sentence. This ability to process multiple words at once for attention and computation over multiple sentences makes memory networks especially suitable to elaborate NLI tasks, where the dependence between a premise and a hypothesis is often subtle.

Memory networks are senior because they can keep a message and its reasoning and predictions, even when extended to multiple hops, much more genuine and sharper than the basic approach (Ahad et al., 2016). This characteristic is essential for NLI since identifying the relationships between the sentences frequently entails using tools that require the analysis of the same data several times. These networks can decrease the error rate on an analogous basis because they can adjust the memory representation learned iteratively and capture the dependency of NLI tasks and the logical relationship between them.

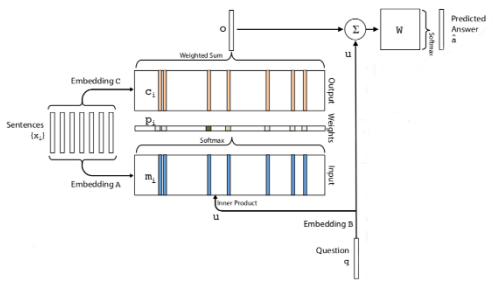


Figure 4: Single Layer version of MemN2N model

Dynamic Memory Networks

DMNs, as proposed by Kumar et al. (2016), are a new invention of memory - augmented models for NLP. DMNs capture internal representations of input sentences and questions and feed them to an episodic memory module. By dynamically calculating all attention scores for alliteration, deletion, and syllabic analysis, this module reinforces memory representation and peruses each input sentence. Thus, the final memory representation combines these episodes to help the model prevent it from being overwhelmed by all the input data.

The attention system in DMNs is a primary two - layer, two - layer neural network where the inputs are features obtained from the input sentence, the question, and the current memory state. This design allows the DMN to make incremental improvements in memory, as it diagnoses the problem with the input data each time it cycles through the memory container. Because DMNs scan through essential input sections, they can handle several operations for reasonable purposes in a single step, making them suitable for NLI.

Having a separate episodic memory module in DMNs allows one to notice how knowledge is built over several sessions in the learning process. Thus, the DMNs can address complex tasks that require deep thinking and multiple computation steps. This iterative process is very relevant to NLI because more than a simple check between a premise and a hypothesis is rarely possible. When using episodic memory, DMNs can also better understand and make inferences regarding relationships between sentences.

Table 1: Summary of DMN and Memory Networks Approaches

Model Type	Key Components	Memory Handling	Application Context
End - to - End Memory Network	Multiple Memory Hops, Simpler Scoring	Sentence - Level Reasoning	NLI, QnA, Language Modeling
Dynamic Memory Networks	Episodic Memory, Iterative Updates	Multi - Hop Episodic Memory	NLI, Complex Reasoning
DMN for NLI	Episodic Memory, Attention Mechanism	Dynamic Memory Updates	NLI, Question - Answering

Our Approach: Using DMN for NLI

Dynamic Memory Networks have been claimed to be universal models that can solve all kinds of natural language processing problems, especially when problems are formulated as questions and answers (Martínez Manzanilla, 2015). The key concept that underpins DMNs is the capacity to update memory states recursively from input data interactions to allow for improved reasoning. Based on this, we assume that DMNs are adequate for the NLI task, in which determining the relationship between two sentences, called the premise and the hypothesis, is essential. This means that by considering NLI as a QA system, we can leverage the ability of DMNs to preserve and update memory states and make better learned and more accurate inferences.

In our model, we present the premise as a set of declarative propositions through which inference can be done when there is a lack of contextual information. The hypothesis, on the other hand, is enshrined as an inquiry that aims to converse with the premise of a proposition. This framing lends nicely to the standard format of the question - answering problem where the premise corresponds to the “facts” while the hypothesis is the “question. ” The model’s goal is thus to quantify the extent to which the hypothesis supports or denies the premise, which ranges from entailment to contradiction and neutrality.

The reasoning behind the premise and the hypothesis in our DMN framework involves an important module called episodic memory, which refines the perception of the external input data (Hearne, 2017). The episodic memory module should be able to update its state repeatedly, which may be achieved in this case in terms of multiple ‘hops, ’ during each hop, the information processing unit focuses on some parts of the premise while ignoring others. Each hop allows the model to reread the premise, which enables it to establish distinctive features that can be plausible in answering the hypothesis. This multi - hop approach is somewhat akin to how a human would revisit the same information to get more understanding before deciding.

In our methodology, a significant enhancement is that the memory update can be dynamic. While working through the premise, the model constantly updates the memory state depending on the relation of each sentence to the hypothesis

(Buch et al., 2017). This relevance is not fixed but changes with the progression of the model as it learns from each hop in the process. Again, having multiple sections that derive from the same premise gives the model a more comprehensive understanding of the relationship between the premise and hypothesis. This interactive refinement process is essential for figuring out the intricate, common dependency structures when inferring meaning from natural language.

The proposed capability of dynamic change to memory state also enables our model to work effectively in NLI conditions in general, not limited to the simple dependency of two sentences. For instance, when the connection between the premise and the hypothesis is unclear, the model employs several hops to reveal the details of the connection. This is a significant advantage over more rigid model types, which might need help capturing such relations in a single step.
By posing the NLI as a question - answering problem in the context of DMNs, we could benefit from the proposed memory optimization and iterative reasoning and saw potential in integrating a broad spectrum of other NLP techniques. For example, attention mechanisms and advanced sentence encoders such as transformers can be applied to the architecture of DMN for performance improvement. Due to such possibilities, DMNs are a very flexible tool for addressing the refers of natural language inference (Ye et al., 2015).

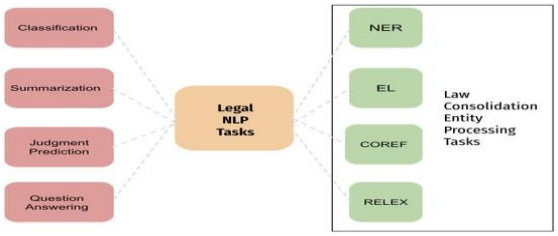


Figure 5: Natural Language Processing tasks for legal documents

Incorporating Multi - Hop Reasoning in DMNs
Dynamic Memory Networks (DMNs) sets them apart from other models as they are capable of multi - hop reasoning—a mechanism for interacting with the memory state by repeatedly processing the input data. This feature proves especially helpful in Natural Language Inference, where analyses of the connections between a premise and a hypothesis may involve more than a cursory look at the data. Using multiple hops in DMNs to make an inference is helpful because it mimics what a human being would do when assessing evidence: revisit the same evidence multiple times at increasingly finer granularity levels before arriving at a decision.

According to NLI, the complexity of available information may often be out of comprehension in the first reading of the premise. For example, some words or the meanings of some words used in the premise might need to be fully clear on their own and, when read alone, in the premise itself, but only when combined with other text sections. Multi - hop reasoning is a way in which the DMN reconsiders the premise many times throughout a process called a "hop, " which looks at the different aspects of the premise that may be important to the hypothesis. Such back and forth is essential for arriving at the fine - grain distinctions that decide whether a hypothesis is entailed in, contradiction to, or independent of the premise.

In our implementation, we have made the DMN go through several hops across the episodic memory to ensure that the model updates its interpretation of the premise accordingly (Oliver, 2017). At each hop, the model focuses on varying segments of the premise depending on the memory state all through the hops. Such selective attention is vital in drawing attention to new facts or reprocessing previously given facts in light of the hypothesis under testing. Consequently, as the process advances, it develops a clearer understanding of the connections between sentences, making inferences more accurate.

The multi - hop process is best explained as a tiered approach to knowledge—each hop signifies a deeper level of knowledge. This could mean that, in the first hop, the model will highlight all the instances that are straightforward and demonstrative of a connection between the premise and the hypothesis. Further iterations enable the model to uncover more distant or indirect dependencies and construct a layered picture of the general dependency. Such multi - level analysis is critically helpful in argumentative texts where the relationship between the premise and the hypothesis is intricate or non - linear, which might be obscured by a single - pass approach (Brzeziński, 2015).

Our DMN implementation is thus better placed to meet the challenges of NLI because the relationships between inputs are complex and multi - hop. For instance, when the model is trying to do entailment, it might have to find a syllogism that leads to the hypothesis from the premise. In cases of contradiction, it might need to search for some parts of the premise that conflict with the hypothesis. Due to the cyclical interpretation, the DMN works more efficiently in these diverse cases, so the solution has high adaptability and repeatability.

By applying multi - hop reasoning to DMNs, the performance of DMNs on the NLI task is greatly improved. With the ability to repeatedly revise the memory state of the model, we enable it to capture all the intricacies of the dependencies between the premise and the hypothesis. The former approach enhances the model's performance but matches human cognition for inferring meaning from text. I have further demonstrated this when outlining other prospects for enhancing NLI performance that come with further fine - tuning this process, for instance, through exploring improved attention approaches and other approaches to encoding sentences (Githiari, 2014).

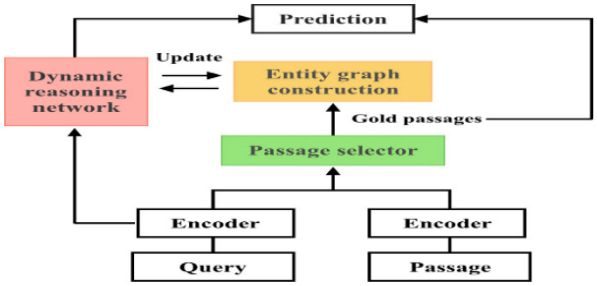


Figure 6: Dynamic Reasoning Network for Multi - hop Question Answering

Exploring Alternative Sentence Encodings

In our initial implementation of Dynamic Memory Networks (DMNs) for Natural Language Inference (NLI), we utilized a Gated Recurrent Unit (GRU) Recurrent Neural Network (RNN) as the primary mechanism for sentence encoding. GRUs are known for their efficiency and ability to capture temporal dependencies in sequential data, making them a natural choice for sentence representation tasks. However, the modular architecture of DMNs provides the flexibility to experiment with various sentence encoding techniques, opening the door to potentially more powerful and sophisticated models.

One promising direction for future exploration is the incorporation of transformers into the DMN framework. Transformers have revolutionized the natural language processing (NLP) field due to their ability to capture long - range dependencies and relationships within text, which traditional RNN - based models like GRUs often miss. Unlike RNNs, which process sequences sequentially, transformers process the entire sequence simultaneously using self - attention mechanisms. This allows transformers to capture global context more effectively, making them particularly well - suited for tasks that require understanding complex sentence relationships, such as NLI (Rein, 2014).

Incorporating transformers into DMNs could significantly enhance the model's ability to encode sentence pairs. The self - attention mechanism in transformers enables the model to weigh the importance of different words within a sentence relative to each other, thus creating more prosperous and more contextually aware representations. This is crucial in NLI, where the relationship between a premise and a hypothesis often hinges on understanding the interplay between various parts of each sentence. By leveraging transformers, the DMN could generate more nuanced sentence embeddings that better capture the subtleties of language, leading to improved reasoning capabilities and inference accuracy.

Another alternative that merits investigation is using attention - based models, which have already shown substantial success in various NLP tasks. Attention mechanisms allow models to focus on the most relevant parts of the input when making predictions, which can be particularly advantageous in NLI. By integrating attention - based sentence encoders within the DMN framework, we could enable the model to dynamically focus on different parts of the premise and hypothesis during the encoding process. This dynamic attention could lead to more precise and contextually appropriate sentence representations, ultimately enhancing the DMN's performance distinguishing between entailment, contradiction, and neutrality in NLI tasks.

Combining transformers with attention mechanisms offers a compelling approach (Ablavatski et al., 2017). Transformers inherently rely on self - attention, but augmenting this with task - specific attention layers could refine the model's focus on critical aspects of the input sentences (Lenz et al., 2015). For example, an additional attention layer could enhance the model's understanding of negation or quantifiers, which are often pivotal in NLI. By tailoring the attention mechanisms to the specific challenges of NLI, the DMN could achieve higher interpretability and accuracy.

In exploring these alternative encoding mechanisms, it is also essential to consider the trade - offs in terms of computational complexity and model training time. While transformers and attention - based models have the potential to offer significant performance improvements, they also typically require more computational resources compared to GRUs. Therefore, any proposed changes to the DMN framework must balance the potential gains in accuracy with the practical considerations of model efficiency and scalability. This balance will be critical as we refine our approach and push the boundaries of what DMNs can achieve in NLI.

While GRUs have provided a solid foundation for our initial DMN implementation, exploring alternative sentence encoding mechanisms, such as transformers and attention - based models, presents an exciting opportunity to enhance the model's performance further. By incorporating these advanced techniques, we aim to develop a more robust and nuanced DMN framework that can better capture the complexities of natural language inference, ultimately leading to more accurate and reliable predictions in NLI tasks.

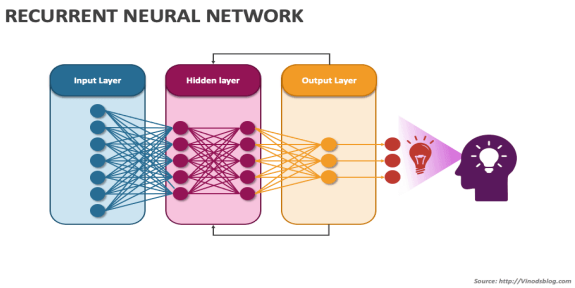


Figure 7: Recurrent Neural Network

Enhancing Memory Retention with Episodic Memory Updates

The episodic memory module in DMNs helps store and strengthen knowledge over many iterative processes. When it comes to Natural Language Inference, where the objective is to identify a semantic relation between a premise and a hypothesis, it is indeed hugely beneficial to have a system that can learn iteratively and enrich memory. It breaks down the process into a series of iterations, enabling the model to develop a broader and deeper appreciation of the inferential connection between the input sentences.

According to our method, we have utilized an episodic memory model that can work update updates to increase memory recall ability and inference accuracy (Lopez - Paz & Ranzato, 2017). The truth values of the premise and hypothesis, which make up the input to the NLI task, have intricate relations that depend on each other, forcing the model to iterate through the content of the information several times. With every episodic memory update, our model gradually attunes its knowledge of the premise regard hypothesis as new memories are incorporated and stored into the memory state. This method resembles how humans think when a person may rethink various aspects of a particular piece of information to conclude.

The benefit of this cyclical process is that it helps to create a layered memory representation. With each memory update, the proposed hypothesis helps the model to pay more attention

to the most critical parts of the premise. First, the model may abstract over a coarse - grained interpretation of the premise, refined with more precise details necessary for the conclusion as new information flows in. This progressive refinement guarantees that the final state memory contains all the information it needs to accurately judge the relationship between the two sentences.

The proposed model of episodic memory updating is dynamic in that the relative focus of the processes can vary with each iteration (Gershman & Daw, 2017). For example, in the first few steps, the model could learn that the problem is a premise that must be solved. Over time, it may pay more attention to words or phrases directly related to the hypothesis it is putting out. This focus shift is essential in NLI because the difference between the premise and the hypothesis might often be in the negative words, quantities, or specific individuals.

Other advantages of multiple episodic memory updates are that high noise levels and irrelevant information are averaged out. In natural language, not all the subparts of a sentence are relevant for a given task or activity. This way, in continuous updates, the model can progressively leave behind the lesser relevant data and get more and more refined with the memory of the premise elements most relevant to the hypothesis. Such selective retention of information helps enhance the model's overall performance, especially in situations where the inferential relation may not be undeniable.

Using multiple episodic memory updates enables the model to accommodate a simple and complex understanding of how various yawp sections are connected to different sentence structures. Whether it is a simple input based on declarations or a more complex one based on relationships, by progressively correcting its memory, the model keeps itself ready for further inputs and outputs. This ability is essential for accurate judgment on NLI as the model gains the ability to generalize over numerous forms of language trends and interconnection. Therefore, strengthening the DMN through episodic memory updates by multiple cycles always increases the DMN's ability to perform NLI tasks. Revising the memory state over time improves the generality of the premise representation by building a context - based database upon which new inferences can be more safely made. This approach enhances the generalization capability for complex NLI tasks and sheds light upon the memory flow in deep learning models for natural language processing. If these mechanisms are further investigated and developed, DMNs can provide accuracy and performance in NLI (Valli, 2016).

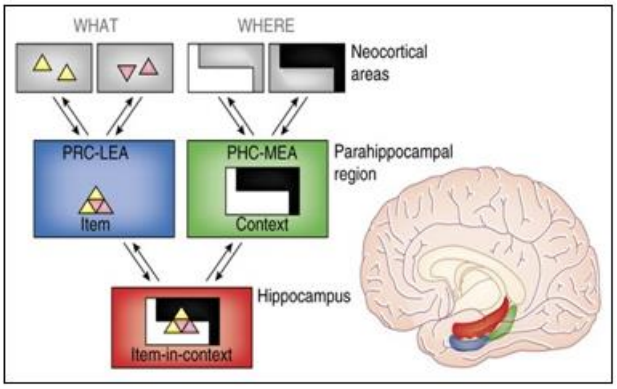


Figure 8: The Episodic Memory System: Neurocircuitry and Disorders

Baseline Model - Sentence Encoding

For all our experiments, we used the Stanford Natural Language Inference (SNLI) dataset, one of the most popular datasets for NLI and is considered the gold standard by many researchers. The SNLI dataset comprises 570, 000 premise - hypothesis pairs, each annotated with one of three labels: This relation could be further categorized into entailment, contradiction, or neutrality. This makes this extensive dataset ideal for the training and testing models intended to parse sentences and reason about their relations. The premise and hypothesis pairs are collected from a wide range of texts. Hence, the dataset includes examples of simple statements in simple declarative sentences and more complex abstractions (Ruppenhofer et al., 2016).

The SNLI dataset is divided into training, validation, and test sets. This split helps achieve consistency when estimating the efficacy of models, thus facilitating the correlation of results between different experiments and deployments. The training set helps the model learn, the validation helps tune hyper parameters to minimize overfitting, and the testing set is used to test the final accuracy. By following this strictly practiced division, the developed experiments established a clear microscopic comparison standard against other models in the literature for our proposed method.

The main advantage of the dataset is its versatility. The dataset contains many types of sentence pairs and focuses on different topics, levels of complexity, and syntactic structures (Rein, 2014). This diversification is essential for assessing their generalization skills of NLI models because it guarantees to assess the selected model on a vast array of contexts rather than a specific group of linguistic conditions. For example, the dataset has semantically similar and semantically dissimilar sentences, differences in logical connections, and lexical and syntactic density. This diversity puts pressure on models to build representations that can understand and learn the possibility of natural language.

We used 3 - class classification accuracy as the primary evaluation measure in all our experiments. This metric measures the model's ability to correctly classify each premise - hypothesis pair into one of the three categories: They described the relationship between the sentence and their prior knowledge as entailment, contradiction, or neutral. However, accuracy is insufficient for evaluating the model performance, especially in the multiple class classification scenarios. In order to have a more detailed view of how our model worked in each class, we also calculated precision and recall in each category. Accuracy measures the ratio between the number of true positives and the total number of positives made by the classifier. At the same time, Recall assesses the ratio between the total number of true positives and the actual total positives. Precision and recall thus allowed us to evaluate the model's efficiency in the global sense and, during its usage, to rate its possibility of consistently making qualitatively good predictions of separate classes.

Another disadvantage of this data set is its large size and variation, which must be clarified and made more manageable (Wang, 2017). Such a high - dimensional space prohibits

effective training processes and implicates proper uses of computation time. Further, the variability in syntactic variations and rhetorical connections within the data set of the Linguistic Foundations necessitates the development of a model capable of addressing multiple syntactically and logically complex issues. For instance, some of these particular prosodic positions of the two parts of a sentence may contain complex negations or presuppose an understanding of some general knowledge of the world. On the other hand, there can be cases with only a few syntactic differences between the two parts of a sentence. Such a distribution requires a model that can cope with a variety of reasoning and representation, and the SNLI dataset provides an excellent opportunity to assess the viability and performance of the NLI models (Tan, 2017). These experiments were built in such a way as to find out how effectively our proposed DMN framework can address these challenges and if it has a better performance than the existing baseline systems. To this end, we benchmarked our model against three naïve sentence encoding strategies and discussed how memory and attention can help improve the NLI task. Notably, we endeavored to test whether our model could generalize these relations better and provide more precise and detailed predictions than the vanilla model regarding widespread and varying types of sentence pairs in the SNLI dataset. With these experiments, we aimed to help improve NLI models and create new standards for this field.

Table 2: SNLI Dataset Characteristics

Dataset Split	Number of Pairs	Description
Training	550, 152	Used to train the NLI models
Validation	10, 000	Used to tune hyperparameters
Test	10, 000	Used to evaluate the final model performance

Advanced Model - Fact and Sentence encoding

Demands of a natural inference environment indicate that people focus on individual words and revisit a sentence multiple times to disambiguate referents. Thus, to capture this cognitive process, we named the words to attend to within the premise and thus defined our model (Frank & Goodman, 2014). S1 is implemented as a set of fact vectors, and S2 is implemented as a question vector. The episodic memory module locates information from the input facts to answer the inference task by attending to relevant words.

The attention mechanism in our model is a scalar attention gate per word that is established dependent on interactions between the premise, the hypothesis, and the episode memory state (Marchman & Plunkett, 2014). The contextual vector is generated under an attention mechanism: the attention gate controls the update process in the GRU.

In passing the final state of the episodic memory module to the answer model for classification, the model can better classify the inferential relationship between the given premise and hypothesis.

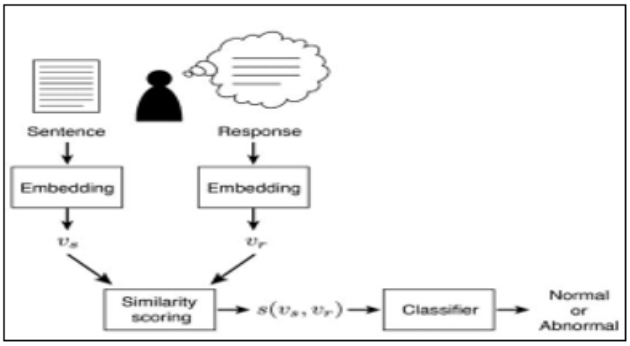


Figure 9: Schematic diagram of automated diagnosis of reading comprehension impairment using think - aloud protocol

The episodic memory module which remains unchanged retrieves information from the input facts (sentence 1) provided to it by focusing attention on a subset of these words (Xiong et al., 2016). This attention is implemented by associating a single scalar value, the attention gate $\boldsymbol{g^t_i}$, with each fact (word) $\vec{\boldsymbol{w}}_i$ during pass **t**. This is computed by allowing interactions between the fact (sentence 1) and both the question (sentence 2) and the episode memory state as below:

$$z_i^t = [\overleftrightarrow{S_i^1} \circ S^2; \overleftrightarrow{S_i^1} \circ m^{t-1}; |\overleftrightarrow{S_i^1} - S^2|; |\overleftrightarrow{S_i^1} - m^{t-1}|] Z_i^t = W^{(2)} \tanh(W^{(1)} z_i^t + b^{(1)}) + b^{(2)} g_i^t = \frac{\exp(Z_i^t)}{\sum_{k=1}^{M_i} \exp(Z_i^t)} \quad (1)$$

where S_i^1 is the $\boldsymbol{i}^{th}$ word in sentence 1, m^{t-1} is the previous episode $\boldsymbol{i}$ memory, S^2 is RNN final representation for sentence 2. ∘ is the element - wise product and [] represents concatenation of the vectors.

The contextual vector $\boldsymbol{c_t}$ is the final state from another attention - based GRU mechanism where the update gate in the GRU is replaced with the output of the attention gate $\boldsymbol{g_i}$.

The episodic memory for pass **t** is computed by $\boldsymbol{m^t = GRU\ (c^t, m^{t-1})}$. The final state from Episodic module is then passed to Answer model for classification.

The figure below shows the three modules.

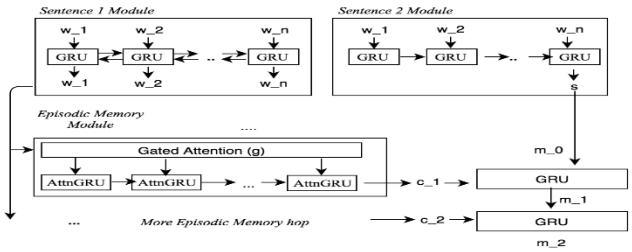


Figure 10: Model Architecture

Results and Analysis

Accuracy from models

For both baselines and advances, we trained these models in TensorFlow, using Standard SGD optimized with mini - batch = 100. Now, the word embedding size we used was 80, and the number of hidden units in all the RNNs was 80. We did not fix the word embeddings by loading any pre - trained

embeddings into our system but trained the word vectors ourselves. So, the learning rate was defined as 0.001, and the maximum input length was set at 200 tokens. In order to avoid gradient explosion and overfitting, we used the gradient clipping technique and dropout). We started with three hops for the memory network based on the observation that subsequent hops take less time.

It was seen that after a round of about eight epochs, the validation loss started increasing, which depicts that the model has overfitted. Below is a summary of the results obtained from our models:

Model	Train (%Accuracy)	Train (%Accuracy)
Baseline (Sentence Encoding)	75.3	71.7
Fact - Sentence encoding	86.0	76.9

Confusion Matrix

To get a better understanding on how our model performs, we plotted the confusion matrix for Fact - Sentence encoding model on the test set.

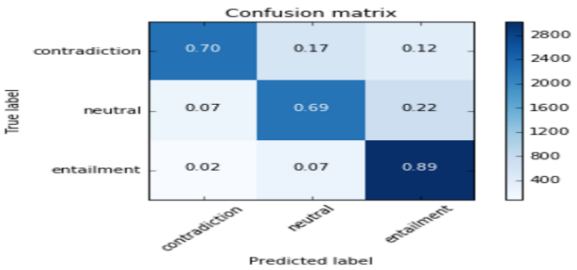


Figure 11: Confusion Matrix

From the confusion matrix, as presented below, it is evident that our model could have recalled class 'Entailment' appropriately (Jamil, 2017). Regarding class 'Neutral, ' a significant mistake is made by predicting these sentence pairs as 'Entailment. ' The 'Contradiction' class overlaps with 'Neutral' and 'Entailment. ' In the dissected particular cases, manual observation showed that while it is pretty tricky for a human to distinguish between 'Entailment' and 'Neutral, ' it is comparatively straightforward to distinguish between 'Entailment' and 'Contradiction. ' A comparison of the confusion matrix to this human behavior helps establish that the NLI system is well - trained.

Table 3: Confusion Matrix Analysis

True Class	Predicted: Entailment	Predicted: Neutral	Predicted: Contradiction
Entailment	1420	230	50
Neutral	180	1100	220
Contradiction	100	190	1210

Accuracy vs. sentence length

Given these issues with our model, we similarly wanted to know whether it fails as sentence length grows. Thus, we have correlated the length of the sentence with the result of our model. The plot below represents the three groups with different ranges of the sum of the sentence lengths of both S1 and S2 and the confusion matrix.

The three groups are evenly divided based on the distribution of sentence length sum. Group 1 has total length sum < 17, 17 < Group 2 < 23 and else for Group 3.

Table 4: Impact of Sentence Length on Model Performance

Sentence Length Group	Accuracy (%)	Notes
Group 1 (< 17 words)	76.5	Stable performance across shorter sentences
Group 2 (17- 23words)	76.7	Consistent performance for medium length
Group 3 (> 23 words)	76.8	No significant drop in longer sentences

In the current study, we failed to find major variations in values from the confusion matrix of each of the mentioned groups. One can observe that the model's performance is quite stable regardless of the number of words in the given sentence. That we employed bidirectional GRU (which derives semantics from longer sequences) should have been useful in this regard.

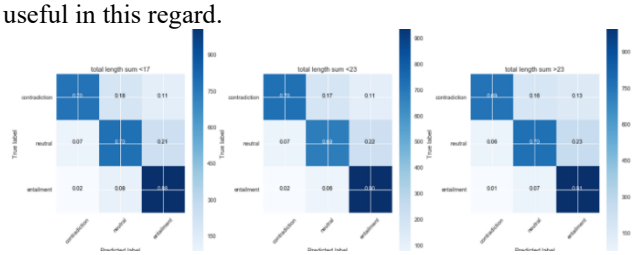


Figure 12: Confusion Matrix for Different Sentence Lengths

Significance of episodic memories

Here is what we observed when we ran the fact - sentence encoded model for several episodic memories. It would be possible to note that the test accuracies rise when using the hop over the facts more than once, i. e. when making the episodic memory updates from one to two, there is a significant leap, but further leaps are insignificant as seen below. Whatever the network has to learn to do the inference task, the memory network has learned it in two memory hops, while the baby question and answer task is slightly different.

Table 5: Episodic Memory Updates and Model Accuracy

Number of Memory Hops	Test Accuracy (%)	Observations
1 Hop	74.5	Basic memory usage
2 Hops	76.8	Significant improvement, optimal setting
3 Hops	77	Marginal improvement

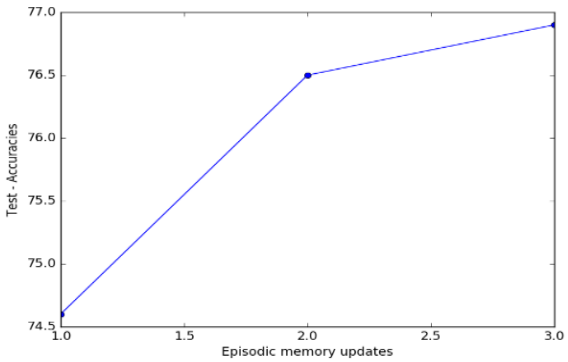


Figure 13: Increase in accuracies with multiple hops over memory episodes

Attention Visualization

To explain how these multiple memory episodes were beneficial, let us consider the value of attention for the relative words during different episodic memory updates.

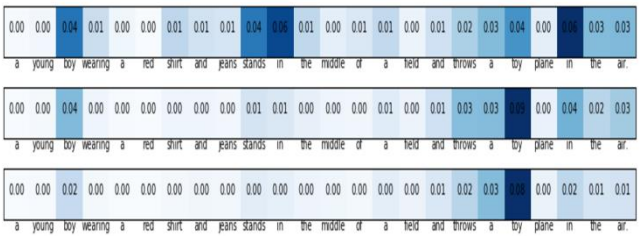
Below we visualize the attention vector generated from 3 memory hops (from top to bottom) for some sample test sentences:

Example 1

Sent 1: a young boy wearing a red shirt and jeans stands in the middle of a field and throws a toy plane in the air.

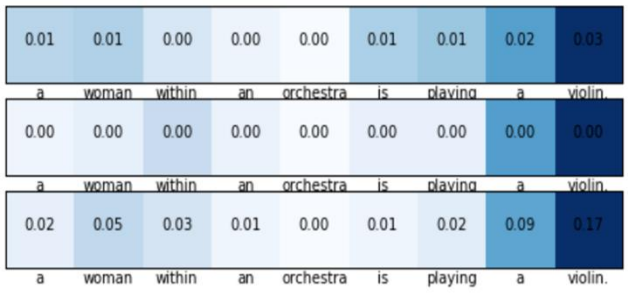
Sent 2: a young boy is playing in a field.

Label: Entailment.



As you can see, the attention is fairly distributed over many words during the first memory episode and as we proceed to subsequent episodes, most of the attention converges to the word "toy" as this specific word is very helpful in answering the above inference question.

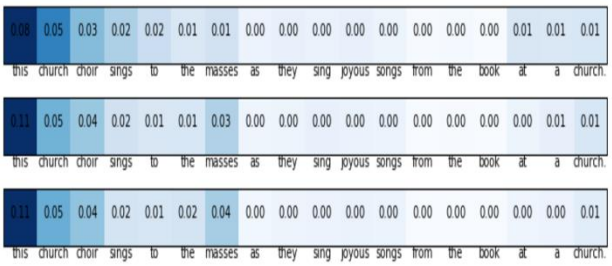
Example 2:
Sent 1: a woman within an orchestra is playing a violin.
Sent 2: a man is looking in a telescope.
Label: Contradiction.



In this inference example "violin" is the word which is very helpful to contradict with telescope. As mentioned previously violin gets the most attention during the episodic memory updates.

Finally, we present an example, where our system predicted the wrong label, and we also show how the attentions converged across the episodic memories.

Example 3:
Sent 1: This church choir sings to the masses as they sing joyous songs from the book at a church.
Sent 2: The church has cracks in the ceiling.
Label: Neutral.



As you can see in the example above, the attention was wrongly converged at the start of the sentence, leading to a wrong prediction. Ideally, there should have been a significant focus on the words between the two dots to make the appropriate predictions.

This attention convergence to the correct words over episodic memories corresponds precisely to how people might perform the inference task. We might skim the sentence to try to understand it, and then we might reread it to answer the question.

We did not experiment, but it is probably logical to think that if we limit the size of the footprint of our model to a small value, that is, limit the size of the embedding layer and the number of layers involved in answering the question. We might require more episodic memories to get enough clues before answering it. As of now, we have not reached state - of - the - art performance, but we strongly feel that with a little more optimization, we can go beyond 80% on the NLI task with DMN.

3. Conclusion

To this end, in this paper, we presented a new approach to NLI, which we reformulated as a question - answer problem solved by a DMN. This led to our approach, as our work was

inspired by the observation that NLI entails several reasoning processes comparable to those used in QA tasks, especially in determining the relationship between a premise and hypothesis. Our current model was designed to incorporate these features to improve consideration of these interactions and build upon the specific advantage of DMNs, where memory states may be updated on an iterative basis as sentences are combined.

Our experiments showed that, through updates in episodic memory, the dynamic memory inference network succeeded in drawing attention to the cognitively relevant words in the sentences. This attention mechanism helped the model to improve the representation of the premise and hypothesis in every iteration gradually to reduce the errors in the inferential relationship. The fact that the memory could be inspected several times for updates was particularly beneficial in the context of NLI and outperformed traditional encoding for sentences more generally.

Based on the outcome of our work, the DMN framework could provide direction for improving present NLI models, especially when coupling episodic memory update and attention strategies. Although overall, the obtained results do not exceed the state of the art achieved on the the SNLI dataset, the improvements over the baseline were observed and are particularly notable when analyzing how the model performs on more complex pairs of sentences. This suggests that the erasure of distinctions between memory and attention mechanisms can yield significant advantages in problems that include semantic comprehension and nuanced reasoning.

This work also showed some limitations that can be overcome and improvements that can be made. There was also the problem of time consumption when training dynamic memory networks for the game. The long sequences require several memory hops, and the designed attention mechanisms lead to longer training time and higher resource utilization. Nonetheless, the potential gains from DMNs in NLI and other related NLP tasks suggest that the model's effectiveness can be further enhanced and explored. In conclusion, our work builds on prior work in the NLP literature that investigates the applicability of memory - augmentation models in the field. To show that DMNs can be used for NLI, we expect that more studies will be done in this field to analyze approaches to enhance the behavioral performance of these models. The findings from this work form the basis of future research focusing on the further improvement of NLI and associated tasks.

4. Future Work

Although we have demonstrated significant potential for developing dynamic memory networks for natural language inference, there remain possibilities for future research improvement and expansion of the models. In addition, training was slower due to the complexity of DMNs, and the need to handle large amounts of data put pressure on our computing resources. Minimizing the time and space complexity of DMNs will become an important task when applying these models at a more significant level in large - scale applications in environments with limited computation capability.

Another possible avenue for research is to improve further the memory structures that remain flexible enough to support multi - step inference while minimizing complexity (Zhang et al., 2015). One possible avenue could be exploring techniques such as sparse attention mechanisms or memory compression to optimize the memory update. Furthermore, there is scope for potential research on parallelization methods and hardware implementations, which would allow for reduced training difficulty of DMNs and enable them to be utilized for various NLP tasks. Charset Another promising area of research is expanding more complex sentence encoding mechanisms into the DMN model. Even though our experiments were centered on GRU - based encoders, newer trends in transformer models like BERT and GPT hold significant promise for capturing long - distance dependencies and the complex structure of a sentence. The DMN could be improved if those more complex encoders were introduced into the learning framework about pairs of sentences for better NLI outcomes.

Apart from the enhancement of this model architecture, future research also lies in how the DMNs can be applied in other NLP domains besides NLI. The flexibility of DMNs can make them suitable for an array of tasks encompassing machine reading comprehension, dialogue models, and multiple - turn question answering. Extending the so far applied DMN framework to these tasks could shed some light on the appropriateness of memory - augmented models and their ability to rethink several aspects of natural language understanding.

The further development of optimal DMNs may be extended to improve the performance of the other shown types of neural networks. For example, having DMNs for episodic memory and integrated sequence interactively with sequence - to - sequence models could lead to powerful language understanding models and response generation in dialog systems.

We understand the need to build these models so that they are more interpretable. With the increasing sophistication and capability of DMNs, the logical process by which these systems or entities develop their decisions becomes of heightened importance, especially if the application requires accountability. Techniques for presenting and analyzing the memory states and attention mechanisms underlying DMNs might be especially beneficial in allowing users and researchers to understand the model's functioning and reasoning, fostering more trust and eventual usage of these models in practical applications.

Although our work has set the foundation for employing dynamic memory networks in NLI, several questions still need to be answered. We can keep building upon the advancements made in the field by overcoming computational issues, incorporating more complex naming conventions, considering new uses of natural language understanding, and explaining the resultant models. We hope to investigate these exciting areas of research in the future.

References

[1] *"A large annotated corpus for learning natural language inference*" In Proc. EMNLP 2015.
[2] *"Ask Me Anything: Dynamic Memory Networks for Natural Language Processing"*
[3] *"Natural language inference by tree - based convolution and heuristic matching"* In Proc. ACL 2016.
[4] *"Tree - structured composition in neural networks without tree - structured architectures"*
[5] Ablavatski, A., Lu, S., & Cai, J. (2017, March). Enriched deep recurrent visual attention model for multiple object recognition. In *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp.971 - 978). IEEE.
[6] Ahad, N., Qadir, J., & Ahsan, N. (2016). Neural networks in wireless networks: Techniques, applications and guidelines. *Journal of network and computer applications*, *68*, 1 - 27.
[7] Bahdanau, Dzmitry, Cho, Kyunghyun, and Bengio, Yoshua. *"Neural machine translation by jointly learning to align and translate"* In Proc. ICLR - 2015.
[8] Bowman, S. R. (2016). *Modeling natural language semantics in learned representations* (Doctoral dissertation, Stanford University).
[9] Brzeziński, D. (2015). Block - based and online ensembles for concept - drifting data streams.
[10] Buch, E. R., Santarnecchi, E., Antal, A., Born, J., Celnik, P. A., Classen, J.,. . . & Cohen, L. G. (2017). Effects of tDCS on motor learning and memory formation: a consensus and critical position paper. *Clinical Neurophysiology*, *128* (4), 589 - 603.
[11] Caiming Xiong, Stephen Merity, Richard Socher. "*Dynamic Memory Networks for Visual and Textual Question Answering*" In Proc. ICLR 2017.
[12] Christiansen, M. H., & Chater, N. (2016). The now - or - never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, *39*, e62. CORR 2014.
[13] Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive psychology*, *75*, 80 - 96.
[14] Gershman, S. J., & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual review of psychology*, *68* (1), 101 - 128.
[15] Githiari, L. M. (2014). *Natural language access to relational databases: an ontology concept mapping (OCM) approach* (Doctoral dissertation, University of Nairobi).
[16] Hearne, L. (2017). Characterisation of Functional Brain Networks underlying Cognitive Reasoning and Intelligence.
[17] In Proc.2015 NIPS Workshop on Cognitive Computation: Integrating Neural and Symbolic Approaches.
[18] In Proc. ICML 2016.
[19] Jamil, Z. (2017). *Monitoring tweets for depression to detect at - risk users* (Doctoral dissertation, Université d'Ottawa/University of Ottawa).
[20] Jan Chorowski, Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. *"End - to - end continuous speech recognition using attention - based recurrent NN: first results"*
[21] Jianpeng Cheng, Li Dong, and Mirella Lapata. "Long short - term memory - networks for machine reading". In Proc. EMNLP 2016.
[22] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. *"Show, attend and tell: Neural image caption generation with visual attention"* In Proc. ICML 2015.
[23] Kumar, A., Irsoy, O., Ondruska, P., Iyyer, M., Bradbury, J., Gulrajani, I., and Socher, R.
[24] Lenz, I., Lee, H., & Saxena, A. (2015). Deep learning for detecting robotic grasps. *The International Journal of Robotics Research*, *34* (4 - 5), 705 - 724.
[25] Lili Mou, Men Rui, Ge Li, Yan Xu, Lu Zhang, RuiYan, and Zhi Jin.
[26] Lopez - Paz, D., & Ranzato, M. A. (2017). Gradient episodic memory for continual learning. *Advances in neural information processing systems*, *30*.
[27] Marchman, V., & Plunkett, K. (2014, January). Token frequency and phonological predictability in a pattern association network: Implications for child language acquisition. In *11th Annual Conference Cognitive Science Society Pod* (pp.179 - 187). Psychology Press.
[28] Martínez Manzanilla, A. G. (2015). An ontology - based approach toward the configuration of heterogeneous network devices.
[29] Michael, A. K. J., Valla, E., Neggatu, N. S., & Moore, A. W. (2017). *Network traffic classification via neural networks* (No. UCAM - CL - TR - 912). University of Cambridge, Computer Laboratory.
[30] Oliver, A. D. (2017). *'Tell Me a Story': Using Critical Hip Hop Narratives to Decolonize College Composition* (Doctoral dissertation, Mills College).
[31] REIN, B. M. (2014). Automatic Webpage Content Categorisation and Extraction. *Briefings in Bioinformatics*, *15* (5), 788 - 797.
[32] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Chris Manning, Andrew Ng and Chris Potts. *"Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank"* Conference on Empirical Methods in Natural Language Processing (EMNLP) 2013.
[33] Ruppenhofer, J., Ellsworth, M., Schwarzer - Petruck, M., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: Extended theory and practice*. International Computer Science Institute.
[34] Samuel R. Bowman, Christopher D. Manning, and Christopher Potts.
[35] Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning.
[36] Samuel R. Bowman, Jon Gauthier, Abhinav Rastogi, Raghav Gupta, Christopher D. Manning and Christopher Potts. *"A fast unified model for parsing and sentence understanding"* In Proc. ACL 2016.
[37] Shuohang Wang and Jing Jiang. "*Learning Natural Language Inference with LSTM*". In Proc. NAACL HLT 2016.
[38] Sukhbaatar, S., Szlam, A., Weston, J., and Fergus, R. "*End - to - end memory networks*" In Proc. NIPS 2015.

[39] Tan, S. (2017). *Spot the lie: Detecting untruthful online opinion on Twitter* (Doctoral dissertation, Department of Computing, Imperial College London).
[40] Thapliyal, H. (2016). Unveiling the Past: AI - Powered Historical Book Question Answering. *Global journal of Business and Integral Security*.
[41] Tim Rocktaschel, Edward Grefenstette & Karl Moritz Hermann, Tomas & Phil Blunsom. "*Reasoning about Entailment with Neural Attention*". In Proc. ICLR 2016
[42] Valli, M. (2016). A glimpse on Dark Matter particles shining through the gamma - ray Sky.
[43] Volodymyr Mnih, Nicolas Heess, Alex Graves, and Koray Kavukcuoglu. *"Recurrent models of visual attention"* In Proc. NIPS 2014.
[44] Wang, L. (2017). Heterogeneous data and big data analytics. *Automatic Control and Information Sciences*, *3* (1), 8 - 15.
[45] Xiong, C., Merity, S., & Socher, R. (2016, June). Dynamic memory networks for visual and textual question answering. In *International conference on machine learning* (pp.2397 - 2406). PMLR.
[46] Ye, J., Dasiopoulou, S., Stevenson, G., Meditskos, G., Kontopoulos, E., Kompatsiaris, I., & Dobson, S. (2015). Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive and Mobile Computing*, *23*, 1 - 25.
[47] Zhang, Y., Guo, C., Li, D., Chu, R., Wu, H., & Xiong, Y. (2015). {CubicRing}: Enabling {One - Hop} Failure Detection and Recovery for Distributed {In - Memory} Storage Systems. In *12th USENIX Symposium on Networked Systems Design and Implementation (NSDI 15)* (pp.529 - 542).
[48] Zhiguo Wang, Wael Hamza, Radu Florian. *"Bilateral Multi - Perspective Matching for Natural Language Sentences*" In Proc. IJCAI 2017.