Prediction of Answer Keywords using Char-RNN

Pratheek I and Joy Paulose
Dept. of Computer Science, Christ University, India

Article Info

Article history:
Received Jan 20, 2018
Revised May 31, 2018
Accepted Nov 16, 2018

Keywords:
Natural Language Processing
Recurrent Neural Network
Long Short Term Memory
Gated Recurrent Units
Neural Networks

ABSTRACT

Generating sequences of characters using a Recurrent Neural Network (RNN) is a tried and tested method for creating unique and context aware words, and is fundamental in Natural Language Processing tasks. These type of Neural Networks can also be used a question-answering system. The main drawback of most of these systems is that they work from a factoid database of information, and when queried about new and current information, the responses are usually bleak. In this paper, the author proposes a novel approach to finding answer keywords from a given body of news text or headline, based on the query that was asked, where the query would be of the nature of current affairs or recent news, with the use of Gated Recurrent Unit (GRU) variant of RNNs. Thus, this ensures that the answers provided are relevant to the content of query that was put forth.

Corresponding Author:
Pratheek I,
Department of Computer Science,
Christ University Hosur Road, Bangalore, Karnataka, India 560029.
Email: pratheek.i@cs.christuniversity.in

1. INTRODUCTION

In the modern world the amount of information that is presented to the user, or even generated is quite overwhelming. Even though there are various applications that facilitate the consolidation of the information, there is a possibility that some information might be missed which may prove to be vital in other scenarios. One way to compress the information while having it retain contextual information is through the use of Recurrent Neural Networks (RNNs)

RNNs are dynamic systems which are able to generate sequences of information, not only in textual domain (1), but also in various other domains, including but not limited to music (2), (3), image (4) among several others. RNNs work on the basic principle that, a certain data is provided to the system, which it processes and trains itself on, and after a certain number of iterations of training, it begins to generate sequences of text, or other information that is provided initially. In theory, this is considered to be good, but when practically using the system one of the main flaws of the system is exposed, which is that as the system generates text, it begins to lose context and thus, the information it subsequently generates is only based on the few previous words it has learnt, but not from the original content from which it started to train.

To overcome this fundamental flaw, the author uses a variant of the standard RNN called as GRU (Gated Recurrent Unit). When compared with a standard RNN, Long Short Term Memory (LSTM) based variant performs better since it is able to retain information, from the original context that it was trained on, and on the newer context that it learns, thus being able to generate a much more diverse and natural sequences of information. A GRU based RNN is considered to be the next step of evolution for LSTMs.

The closest implementation of this nature was attempted by the Facebook Research team(5). In their paper, they are using a large data corpus to predict answers for factual questions. The system proposed in this paper attempts to provide answers to questions whose answers are considered to be currently 'in news', and may not necessarily have an encyclopaedic entry.

Section 2. contains the Literature Review of the papers that are used as reference material for this paper. Section 3.1. consists of the GRU based RNN model that is used in this paper. Section 3.2. includes the proposed method, a model, and its evaluation. The results are recorded in section 4. Section 5. comprises the future scope and conclusion.

2. RELATED WORK

2.1. Smart Reply: Automated Response Suggestion for Email

Replying to an Email is sometimes a tricky situation, especially with choosing the right words to convey the right meaning which is somewhat a tedious task. In the past there have been many automated response systems built within email clients but most of those are just static sentences that wouldn't be flexible to a given scenario. They would just be simple sentences. In this paper, The authors of the paper, devise a system that would provide an automated response to a received email, based on the content of the email.

In this system, the authors create a sequence-to-sequence long short term memory network to predict the sequences of text. The input is an incoming message, and the output is a series of responses that are generated based on the provided text corpus. The LSTMs were first originally applied to Machine Translation but has since seen success in other domains such as image captioning and speech recognition. (6)

2.2. Improving Context Aware Language Models

In the current age, many automated textual systems use the LSTM model to generate text responses. The authors of this paper propose an alternative approach to the standard RNN used for text generation, which is Recurrent Neural Network Language Model (RNNLM). What they propose is, to replace the domains which are used to provide a sort of an inadequate context for the text generation, with context based variables. These context variables are used to describe certain aspects of the language, such as topic, time, or other language. These context variables are then dynamically combined to create to create a more coherent text. The combination of the context variables with context embedding creates the aforementioned RNNLM.

The data used to test the model is obtained from reddit, twitter and SCOTUS. On the reddit data, the testing is done based on a given comment, and the objective for the model is to identify the subreddit (discussion group of one topic), from which the sentence might have originated. There are eight specific subreddits, and nine general subreddits, and the results show that the RNNLM model is comparatively better on specific subreddits, compared to general subreddits, thus showing that context based language models were much more effective. (7)

2.3. Contextual LSTM (CLSTM) models for Large scale NLP tasks

Text presented in any format is usually in the form of phrases, sentences, paragraphs, sections among others. These formats are a way of abstracting sentences into a combined entity to present a cohesive meaning. In this paper, the authors create a Contextual LSTM (CLSTM) model, and it is tested against a normal LSTM model on Natural Language Processing tasks, such as word selection, next word prediction, and next sentence prediction. The data used on this model was from English Wikipedia, and English Google News site.

The results on the English Google News data, showed improvements across the board when comparing LSTM with CLSTM. In terms of next word prediction, LSTM using words as features had the perplexity of 37, and the CLSTM improvement was about 2 percent. In the case of next sentence selection, the improvement was about 39 percent for the LSTM, when compared with CLSTM, which had an accuracy of 46 percent and, finally next sentence topic prediction in LSTM, using current sentence topic as feature, the perplexity was about 5, and CLSTM improved on it by 9 percent. Thus one can see that the usage of context along with an LSTM has an improving effect over a standard LSTM, for natural language processing tasks. (8)

2.4. Context-aware Natural Language Generation with Recurrent Neural Networks

Natural Language generation is a useful system in various applications, such as response generation in messaging systems, text summarization and image captioning. Most of the natural language generation works only with the provided amount of content, while ignoring the contextual information that is present. However, in real life scenarios, these systems ignore the contextual component, and only focus on the current data to generate text. The authors of the paper, propose two approaches for this. The first is a C2S (context to sequences) model, which encodes a set of contexts into a continuous representation and then decode the representation into a text sequence, through a recurrent neural network. However not all the words may depend on the contexts, some of which may only depend on their preceding words. To resolve this, a gating mechanism is introduced to control when the information from the contexts are accessed. The second model that they propose, is called gC2S (Gated Contexts to Sequences).

The data used for this is customer reviews from sites such as Amazon and TripAdvisor (Travel website, where travelers can review places). In Amazon the reviews were focused from the book, electronic, and movie categories; and for TripAdvisor, the category was hotels. Upon testing, the authors found that, the gC2S model significantly outperforms the C2S model as it adds skip-connections between the context representations and the words in the sequences, allowing the information from the contexts to be able to directly affect the generation of words. (9)

2.5. Generating Sentences from a Continuous Space

The standard Recurrent Neural Network Language Model (RNNLM) generates a single word at a time and does not work from an explicit global sentence variational autoencoder generative model that incorporates distributed latent representations of entire sentences. This factorization allows it to explicitly model holistic properties of sentences such as style, topic, and high-level syntactic features. The RNNLM is a probabilistic model with no significant assumptions since it generates sentences word-by-word based on an evolving distributed state representation. However, when the RNNLM model is broken down into a series of next-step predictions, it does not expose interpretable representation of global features. To overcome such short comings, the authors of the papers propose a concept, which they refer to as variational autoencoder, through which they hope to capture the global features in a continuous latent variable.

A standard RNN predicts words based on conditioning of the previous word and an evolving hidden state. Although effective, it does not learn about the vector representation of the full sentence. In order to represent a continuous latent sentence, one should first find a method to map sentences and distributed representations that can be trained in an unsupervised setting. Sequence autoencoders are somewhat successful, in generating complete documents, an autoencoder consists of an encoder function and a probabilistic decoder model, in the case of a sequence autoencoder, both encoder and decoder are RNNs and examples are token sequences. The Variational Autoencoder (VAE) is a generative model that is based on a regularized version of the standard autoencoder. This model imposes a prior distribution which enforces a regular geometry and makes it possible to draw proper samples from the model using ancestral sampling.

The dataset used for this is a collection of text from K12 e-books, which contains approximately 80 million sentences. One of the test was to see the imputation of missing words and the authors observed that a standard RNN can only predict from a previous set of provided tokens where as the RNN with VAE is much more effective and can provide an intuitive based completion which fit the context much better and provided a more cohesive meaning to a sentence. (10)

2.6. A Neural Attention Model for Abstractive Sentence Summarization

Summarization is one of the most important components in the current world, it is especially required by the news outlets and discussion forums, among others to summarize contents that are otherwise too long to read by users who would not have the time to do so. The objective is to create a condensed form of the text corpus that is able to capture the whole meaning of the corpus. Most common form of summarization is extraction style, where certain parts of the text is extracted out and then stitched together. The other form is abstractive summarization where the approach is to do the summarization from the bottom up. The authors of this paper propose a new concept of summarization called as attention based summarization. It incorporates linguistic structure when compared with other approaches, and can scale easily to train on large amounts of data. Since the system makes no assumptions about the vocabulary of the generated summary, it can be trained directly on any document pairs, and thus can be used to summarize a given data into a headline.

The dataset used by the authors was a collection of 500 news articles from The New York Times and Associated Press, paired with 4 different human generated reference summaries. The authors find that, the proposed model is able to perform comparatively much better than the older approaches and in a more abstractive summaries. (11)

2.7. Neural Turing Machines

The neural network is an excellent concept. It is able to mimic the human brain quite well. But one of the missing components is the memory part of it, the neural network is not able to remember the activity it has completed. It trains for a given task, and executes the task but once completed, it forgets the method, and if the said task is to be performed again, it would have to be trained yet again and then the task could be performed. The authors of this paper propose a concept in which the neural network (in this case, an LSTM) is paired with a memory bank. This neural network is able to interact with both standard input and output vectors, and also with a memory bank.

The authors train the model in performing two simple tasks; reading and writing, and then store the concept of the task into its memory banks. Then they perform a few tests, first being the copy test. The model is provided with random values of 8-bit binary values of random lengths, and then it is to output a value, that is a copy of the initial sequence. The authors find that after training both the LSTM model and an NTM model, the NTM model is able to perform subsequent tasks of copying in much lesser time, and it need not be trained again. Another test that the authors performed was the sorting. The model was provided with random binary values, along with priority which is between -1 and 1. They found that the NTM was able to understand the concept of sorting and was able to perform the task much more cohesively. It was able to approach this concept as a human mind would, compared to an LSTM model. (12)

After examining various papers which are aforementioned, it was found that the system that the authors were looking for does not exist, and thus, through this paper the authors wants to propose such a system.

3. PROPOSED WORK

3.1. RNN Model

In this section, the authors describe the model that is being used in this paper. An example of this RNN Model is shown in Figure 1. The first thing to note is the input which is a vector

$$\mathbf{X} = (x_1, \ldots, x_T) \tag{1}$$

This along with certain weights is passed as input, to a stack of recurrently connected and hidden layers, represented by **A**, to calculate the first hidden vector sequence of

$$\mathbf{h}^n = (h_1^n, \ldots, h_T^n) \tag{2}$$

and the output vector sequence is

$$\mathbf{y} = (y_1, \ldots, y_T) \tag{3}$$

This output vector $\mathbf{y}$ is parameterized, looped and redirected back into the network as a set of inputs. This looping feature is what provides the neural network, its key recurrence nature. To observe the actual nature of the RNN one can see that the ‘unrolled’ RNN is just chained normal neural network, which processes the data as required and passes the information to the next successor.

Now, this can be considered as a single layer network. To make it a ‘deep’ RNN, one can stack them one on top of the other. Now, one of the main problems that occurs as the information traverses the layers is that, the components in the subsequent layers are learning faster compared to the previous layers. This problem is referred to as the ‘Vanishing Gradient Problem’ (13). Here, gradients are the values that refer to speed at which networks are able to pick up and learn the information. This is an important factor in the case of RNNs since the data is propagated back into the network and if the values keep reducing as they propagate through the layers, then eventually it would tend to zero, and that would make the model useless. Thus to overcome this, the output is directly repropagated to input layers, and not to the middle layers.

Now as this process is happening, the other problem is that, as it learns new information, the system fails to keep track of the previously learnt information. To overcome this caveat, an LSTM style of RNN is used. LSTM (14). The LSTM variant of RNN essentially eliminates the main drawback of the standard RNN, which is its ability to remember, and this variant works on a large class of problems.

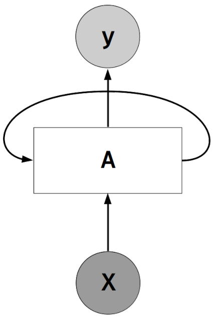


Figure 1. Recurrent Neural Network

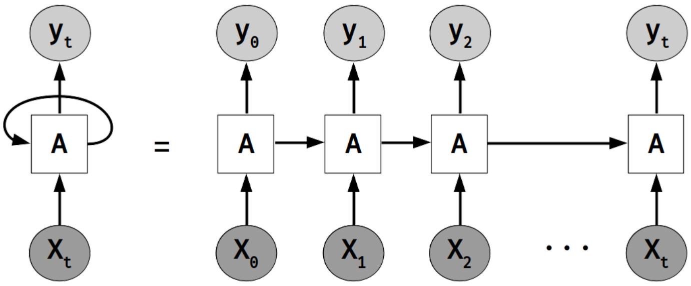


Figure 2. RNN expanded

3.2. Proposed Work

Many algorithms exist to predict keywords, but most of such systems exist for data which is quite large. The authors want to design a system that has a low-overhead, in terms of model building or the amount of data required to predict. The model the authors have chosen is an GRU variant of the Recurrent Neural Network. The authors split the overall process as three subprocesses and then expand each of those subprocesses' individually as shown in Figure 3

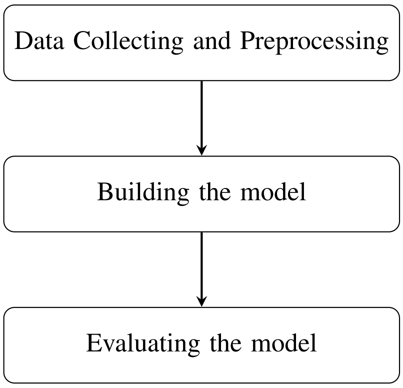
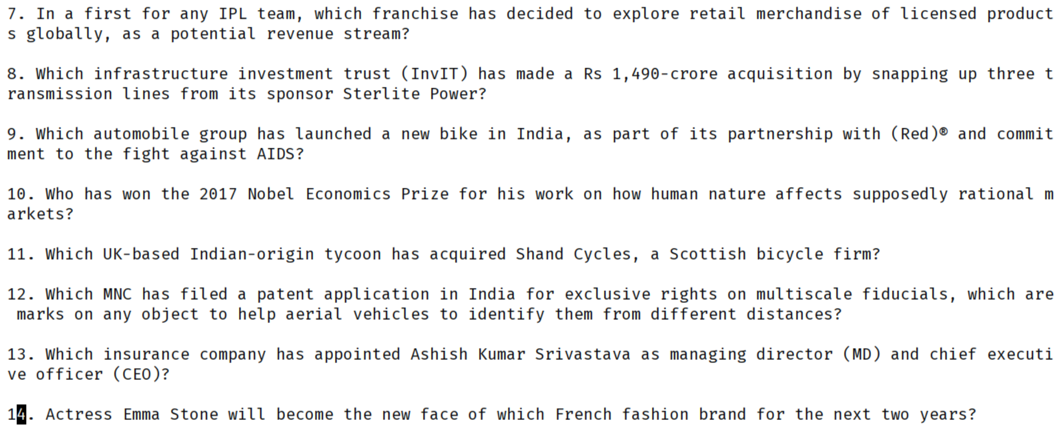


Figure 3. Overall process

3.3. Sub Process 1: Data Collecting and Pre-processing

For this, the data is of homemade or self-made dataset. It is a collection of questions, stored in a text file, interleaved with a new line. An example of the dataset is shown in Figure 4.



```
7. In a first for any IPL team, which franchise has decided to explore retail merchandise of licensed product
s globally, as a potential revenue stream?

8. Which infrastructure investment trust (InvIT) has made a Rs 1,490-crore acquisition by snapping up three t
ransmission lines from its sponsor Sterlite Power?

9. Which automobile group has launched a new bike in India, as part of its partnership with (Red)® and commit
ment to the fight against AIDS?

10. Who has won the 2017 Nobel Economics Prize for his work on how human nature affects supposedly rational m
arkets?

11. Which UK-based Indian-origin tycoon has acquired Shand Cycles, a Scottish bicycle firm?

12. Which MNC has filed a patent application in India for exclusive rights on multiscale fiducials, which are
 marks on any object to help aerial vehicles to identify them from different distances?

13. Which insurance company has appointed Ashish Kumar Srivastava as managing director (MD) and chief executi
ve officer (CEO)?

14. Actress Emma Stone will become the new face of which French fashion brand for the next two years?
```

Figure 4. Example of the homemade dataset

First, the question is searched on a normal internet search engine, and then the results obtained are processed. This is done by first converting all of the content to lowercase, and then the common stop words are removed. For this task, one uses a certain application called SpaCy(15), and NLTK(16), (17) which is used for natural language processing, to clear out the said words and the punctuation marks as well (While, these punctuation marks are useful for a normal human reader,they only add noise to the model). Then, depending on the type of the question, the user can also filter out numeric values which may be present (Having numeric values is useful, only if the answer keywords that are being predicted have them, else they add noise to our model). Once this pre-processing is repeated for all the fetched data, it is stored into a file, and this becomes our input data to the model. The process for which is shown below, Figure 5.

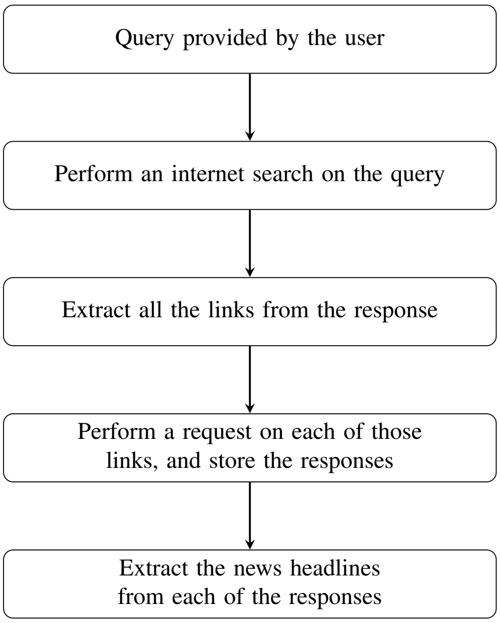

Query provided by the user
Perform an internet search on the query
Extract all the links from the response
Perform a request on each of those links, and store the responses
Extract the news headlines from each of the responses

Figure 5. Data pre-processing flow diagram

3.4. Sub Process 2: Building the model

Here, the input data is split into chunks of certain size. These chunks form the input vectors for the model. Each chunk forms a single input vector and each value from the vector is provided as an input for the first layer of the RNN. So, logically, once these values are “encoded”, the user can start to train the model on these values. After they are processed by the first layer, the output is captured. This becomes the first state. This state is retropropagated as input, back into the model. As the model goes over a new encoded vector each time, it remembers what it has trained on as well as the newer information that it has learnt.

Suppose, the first input is considered to have an error and the user isn’t allowed to modify it, then the magnitude of the error multiplies as the model builds and it causes problems. But, with an RNN one can tweak the data that is propagated from the subsequent layers, thus being able to diminish the error rate in the system. This is because, one can control the amount of information that is being retropropagated back into the system. So, if the first input contains error, then one can take the output after the first iteration, tweak that output and then provide that as the input for the second iteration. So, using the dataset once can train the whole model, and after this one can start evaluating the model.

3.5. Sub Process 3: Evaluating the model

There are two possible ways to evaluate the model. The first method is the manual method, where in the authors manually verify the predictions, by noting the answer keywords and then, manually searching for the query and verifying the results. This method is a bit tedious. The other method is where one can have a program that automates this. This would make it much simpler. In the case of automation, the process would be similar to the manual process. The authors provide a set of questions at a time, five questions, for example. Now, they perform the operation for each of the question, and capture the output. Then, they tokenize the captured output, and compare this with a set answer keywords, that the authors themselves have provided. If the provided answer keywords is said to be found in the tokenized words, then the model is said to have predicted the answer correctly, else the model has predicted inaccurately.

4. RESULTS

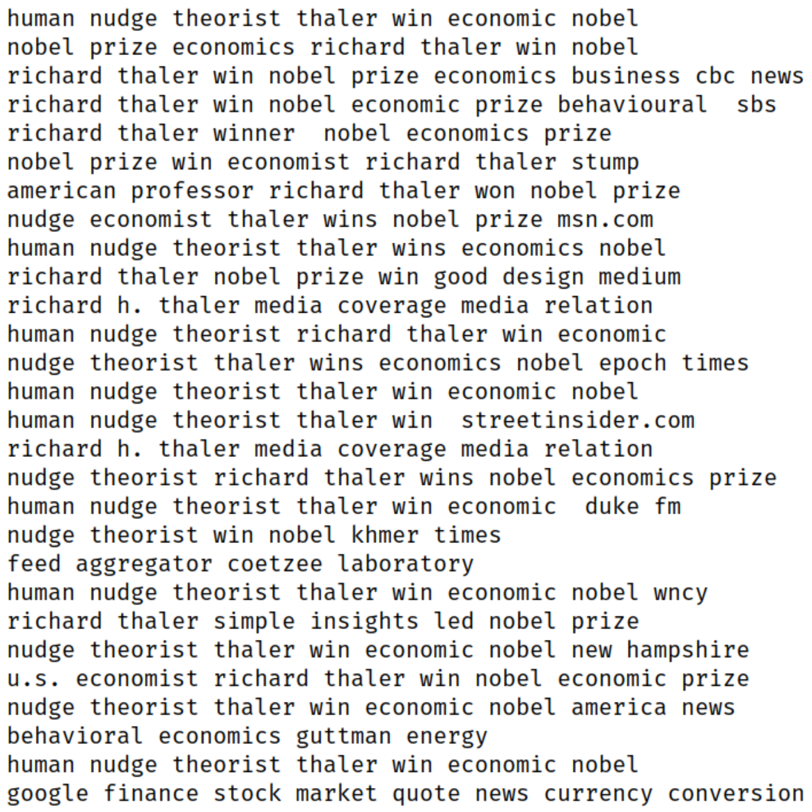
As mentioned in the section 3.5., for the automated process, one has obtained a set of questions from the user, and provided this as the testing dataset for the model. After each complete run of the model, the output was captured and compared with the answers that was provided by the users themselves.

One of the ways in which the system was tested, was by manually providing questions. Here the authors consider two questions, the first being, “`Who won the 2017 Nobel prize for economics?`” and the second question being, “`Which company has launched the Indian Pale Ale (IPA), its fifth beer in India`”.

For the first question the obtained results are shown in Figure 6. After providing this as the dataset for the model to train, the system is able to predict certain sequence of words which would match the results that would be obtained from an internet search query. This is shown in the following figures. Answer keyword patterns predicted by the system is shown in Figure 7 and the answers obtained from internet search is shown in Figure 8.

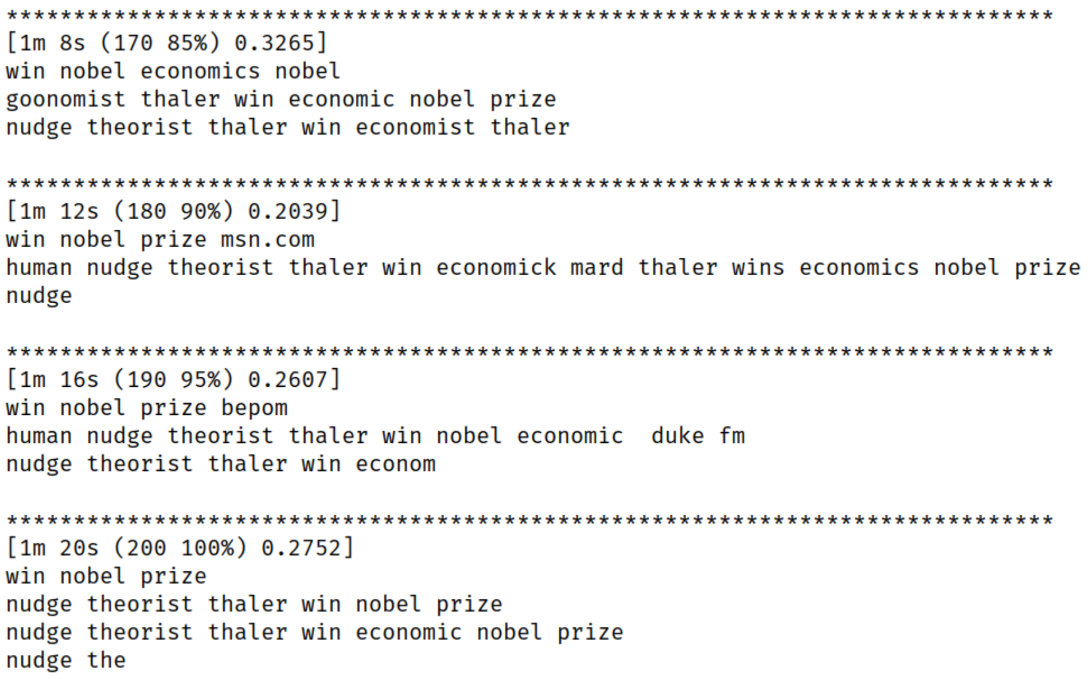
The results of the second question was also verified in a similar manner. The results dataset is as shown in the Figure 9. This is provided as input for the system to train on, and once trained the system begins to generate answer keywords patterns, similar to the previous example and is shown in the Figure 10. The observation to note here is that, when queried via the internet search engines, similar results are yielded, as shown in the Figure11

The observed training loss for each of the questions is shown in the following figures. The graph of the training loss for the first question is shown in Figure 12 and similarly for the second question in Figure 13. One general observation to note in both the cases is that for each iteration, the amount of text being processed increases, and the training loss decreases; indicating that the model is getting more effective in predicting keywords.



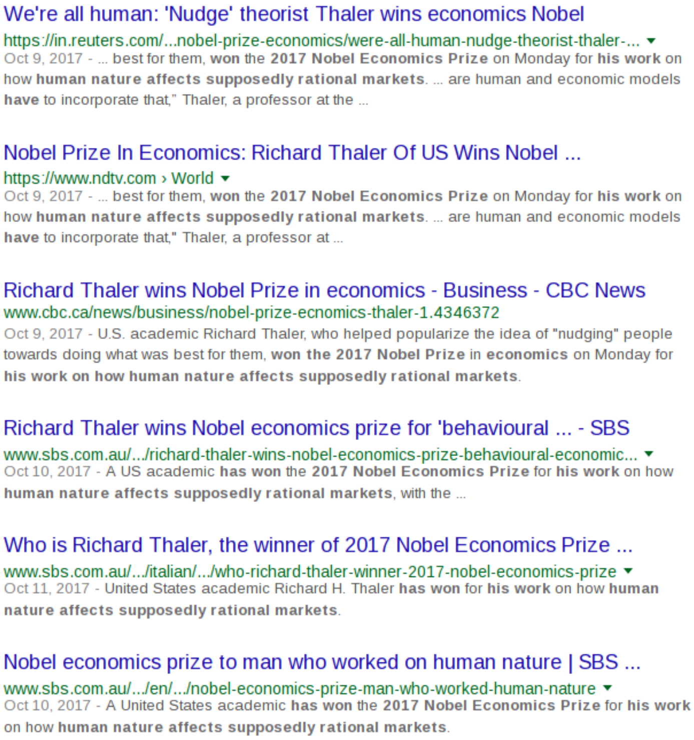
```
human nudge theorist thaler win economic nobel
nobel prize economics richard thaler win nobel
richard thaler win nobel prize economics business cbc news
richard thaler win nobel economic prize behavioural  sbs
richard thaler winner  nobel economics prize
nobel prize win economist richard thaler stump
american professor richard thaler won nobel prize
nudge economist thaler wins nobel prize msn.com
human nudge theorist thaler wins economics nobel
richard thaler nobel prize win good design medium
richard h. thaler media coverage media relation
human nudge theorist richard thaler win economic
nudge theorist thaler wins economics nobel epoch times
human nudge theorist thaler win economic nobel
human nudge theorist thaler win  streetinsider.com
richard h. thaler media coverage media relation
nudge theorist richard thaler wins nobel economics prize
human nudge theorist thaler win economic  duke fm
nudge theorist win nobel khmer times
feed aggregator coetzee laboratory
human nudge theorist thaler win economic nobel wncy
richard thaler simple insights led nobel prize
nudge theorist thaler win economic nobel new hampshire
u.s. economist richard thaler win nobel economic prize
nudge theorist thaler win economic nobel america news
behavioral economics guttman energy
human nudge theorist thaler win economic nobel
google finance stock market quote news currency conversion
```

Figure 6. Example of the fetch results, for the provided query



```
******************************************************************************
[1m 8s (170 85%) 0.3265]
win nobel economics nobel
goonomist thaler win economic nobel prize
nudge theorist thaler win economist thaler

******************************************************************************
[1m 12s (180 90%) 0.2039]
win nobel prize msn.com
human nudge theorist thaler win economick mard thaler wins economics nobel prize
nudge

******************************************************************************
[1m 16s (190 95%) 0.2607]
win nobel prize bepom
human nudge theorist thaler win nobel economic  duke fm
nudge theorist thaler win econom

******************************************************************************
[1m 20s (200 100%) 0.2752]
win nobel prize
nudge theorist thaler win nobel prize
nudge theorist thaler win economic nobel prize
nudge the
```

Figure 7. Answer keyword patterns predicted by the system



We're all human: 'Nudge' theorist Thaler wins economics Nobel
https://in.reuters.com/...nobel-prize-economics/were-all-human-nudge-theorist-thaler-... ▾
Oct 9, 2017 - ... best for them, **won** the **2017 Nobel Economics Prize** on Monday for **his work** on how **human nature affects supposedly rational markets**. ... are human and economic models **have** to incorporate that," Thaler, a professor at the ...

Nobel Prize In Economics: Richard Thaler Of US Wins Nobel ...
https://www.ndtv.com › World ▾
Oct 9, 2017 - ... best for them, **won** the **2017 Nobel Economics Prize** on Monday for **his work** on how **human nature affects supposedly rational markets**. ... are human and economic models **have** to incorporate that," Thaler, a professor at ...

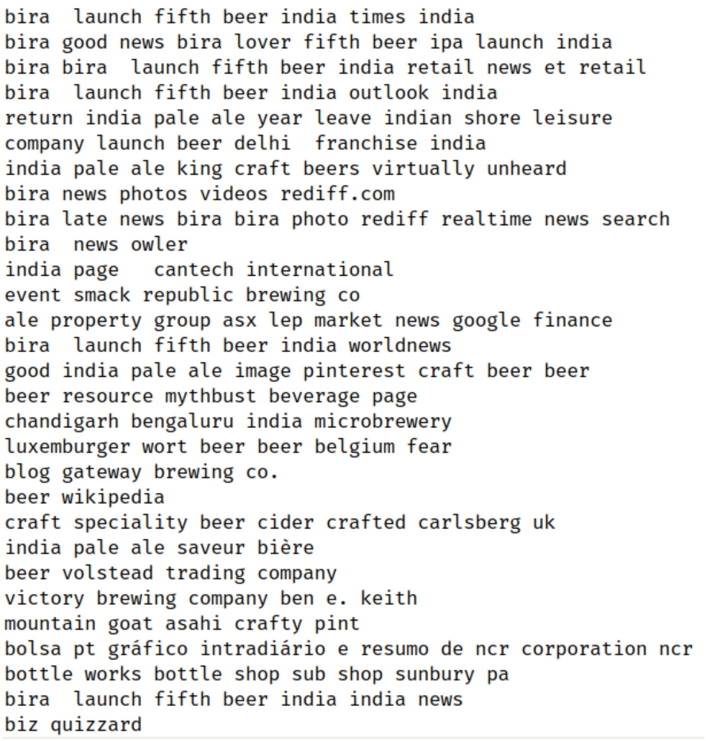
Richard Thaler wins Nobel Prize in economics - Business - CBC News
www.cbc.ca/news/business/nobel-prize-ecnomics-thaler-1.4346372
Oct 9, 2017 - U.S. academic Richard Thaler, who helped popularize the idea of "nudging" people towards doing what was best for them, **won the 2017 Nobel Prize** in **economics** on Monday for **his work on how human nature affects supposedly rational markets**.

Richard Thaler wins Nobel economics prize for 'behavioural ... - SBS
www.sbs.com.au/.../richard-thaler-wins-nobel-economics-prize-behavioural-economic... ▾
Oct 10, 2017 - A US academic **has won** the **2017 Nobel Economics Prize** for **his work** on how **human nature affects supposedly rational markets**, with the ...

Who is Richard Thaler, the winner of 2017 Nobel Economics Prize ...
www.sbs.com.au/.../italian/.../who-richard-thaler-winner-2017-nobel-economics-prize ▾
Oct 11, 2017 - United States academic Richard H. Thaler **has won** for **his work** on how **human nature affects supposedly rational markets**.

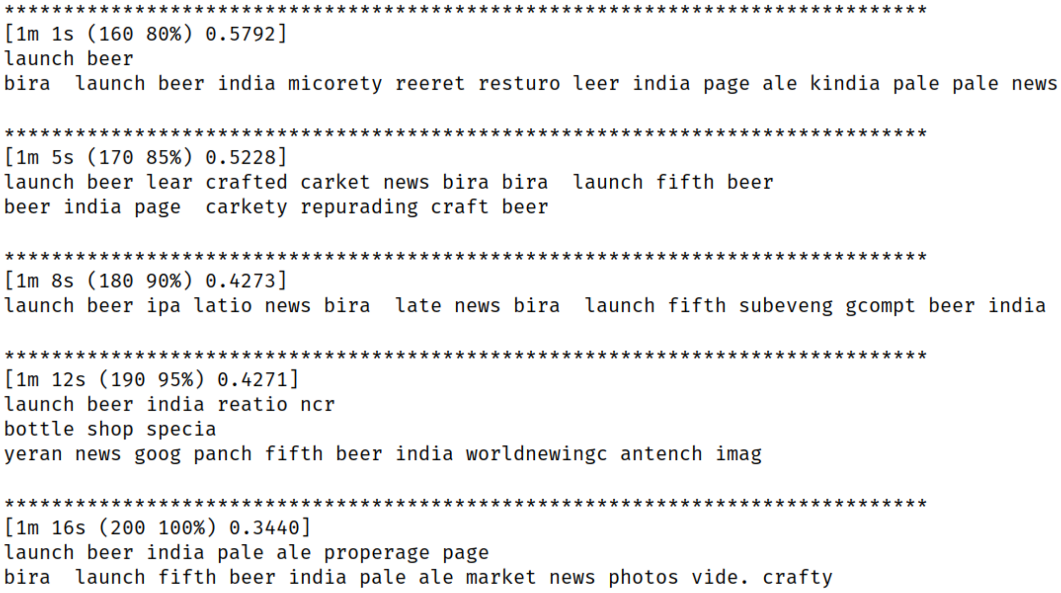
Nobel economics prize to man who worked on human nature | SBS ...
www.sbs.com.au/.../en/.../nobel-economics-prize-man-who-worked-human-nature ▾
Oct 10, 2017 - A United States academic **has won** the **2017 Nobel Economics Prize** for **his work** on how **human nature affects supposedly rational markets**.

Figure 8. Answers as obtained from an Internet search



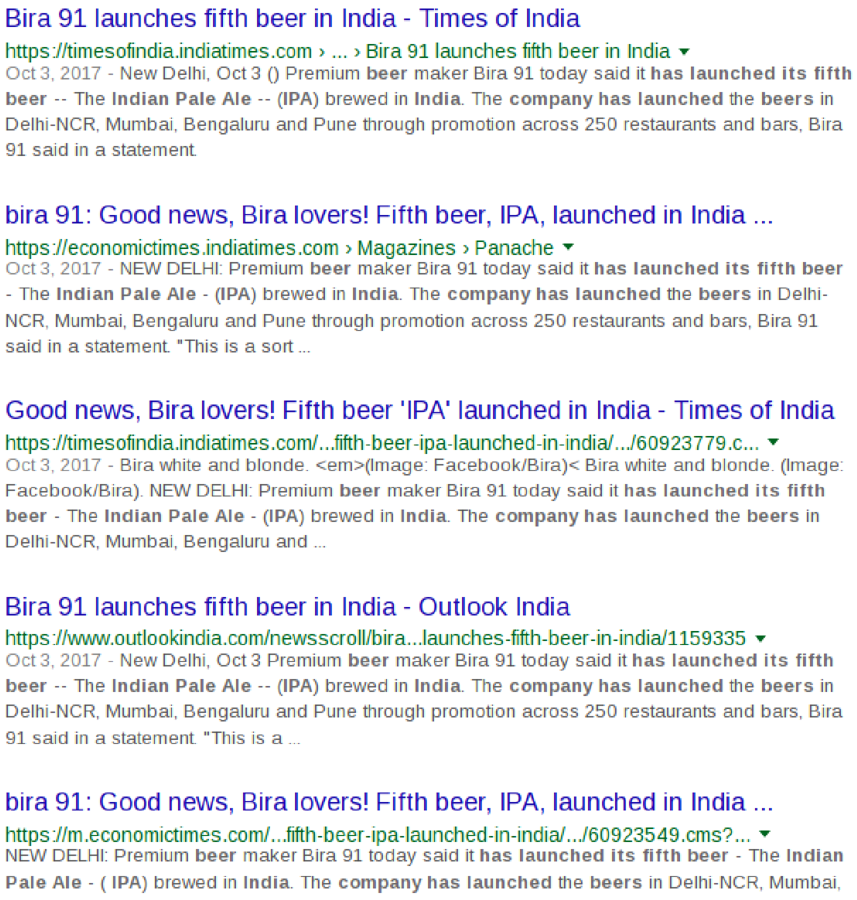
```
bira  launch fifth beer india times india
bira good news bira lover fifth beer ipa launch india
bira bira  launch fifth beer india retail news et retail
bira  launch fifth beer india outlook india
return india pale ale year leave indian shore leisure
company launch beer delhi  franchise india
india pale ale king craft beers virtually unheard
bira news photos videos rediff.com
bira late news bira bira photo rediff realtime news search
bira  news owler
india page   cantech international
event smack republic brewing co
ale property group asx lep market news google finance
bira  launch fifth beer india worldnews
good india pale ale image pinterest craft beer beer
beer resource mythbust beverage page
chandigarh bengaluru india microbrewery
luxemburger wort beer beer belgium fear
blog gateway brewing co.
beer wikipedia
craft speciality beer cider crafted carlsberg uk
india pale ale saveur bière
beer volstead trading company
victory brewing company ben e. keith
mountain goat asahi crafty pint
bolsa pt gráfico intradiário e resumo de ncr corporation ncr
bottle works bottle shop sub shop sunbury pa
bira  launch fifth beer india india news
biz quizzard
```

Figure 9. Example of the fetch results for the second question



```
******************************************************************************
[1m 1s (160 80%) 0.5792]
launch beer
bira  launch beer india micorety reeret resturo leer india page ale kindia pale pale news

******************************************************************************
[1m 5s (170 85%) 0.5228]
launch beer lear crafted carket news bira bira  launch fifth beer
beer india page  carkety repurading craft beer

******************************************************************************
[1m 8s (180 90%) 0.4273]
launch beer ipa latio news bira  late news bira  launch fifth subeveng gcompt beer india

******************************************************************************
[1m 12s (190 95%) 0.4271]
launch beer india reatio ncr
bottle shop specia
yeran news goog panch fifth beer india worldnewingc antench imag

******************************************************************************
[1m 16s (200 100%) 0.3440]
launch beer india pale ale properage page
bira  launch fifth beer india pale ale market news photos vide. crafty
```

Figure 10. Answer keywords as predicted by the system



Bira 91 launches fifth beer in India - Times of India
https://timesofindia.indiatimes.com › ... › Bira 91 launches fifth beer in India ▾
Oct 3, 2017 - New Delhi, Oct 3 () Premium **beer** maker Bira 91 today said it **has launched its fifth beer** -- The **Indian Pale Ale** -- (**IPA**) brewed in **India**. The **company has launched** the **beers** in Delhi-NCR, Mumbai, Bengaluru and Pune through promotion across 250 restaurants and bars, Bira 91 said in a statement.

bira 91: Good news, Bira lovers! Fifth beer, IPA, launched in India ...
https://economictimes.indiatimes.com › Magazines › Panache ▾
Oct 3, 2017 - NEW DELHI: Premium **beer** maker Bira 91 today said it **has launched its fifth beer** - The **Indian Pale Ale** - (**IPA**) brewed in **India**. The **company has launched** the **beers** in Delhi-NCR, Mumbai, Bengaluru and Pune through promotion across 250 restaurants and bars, Bira 91 said in a statement. "This is a sort ...

Good news, Bira lovers! Fifth beer 'IPA' launched in India - Times of India
https://timesofindia.indiatimes.com/...fifth-beer-ipa-launched-in-india/.../60923779.c... ▾
Oct 3, 2017 - Bira white and blonde. <em>(Image: Facebook/Bira)< Bira white and blonde. (Image: Facebook/Bira). NEW DELHI: Premium **beer** maker Bira 91 today said it **has launched its fifth beer** - The **Indian Pale Ale** - (**IPA**) brewed in **India**. The **company has launched** the **beers** in Delhi-NCR, Mumbai, Bengaluru and ...

Bira 91 launches fifth beer in India - Outlook India
https://www.outlookindia.com/newsscroll/bira...launches-fifth-beer-in-india/1159335 ▾
Oct 3, 2017 - New Delhi, Oct 3 Premium **beer** maker Bira 91 today said it **has launched its fifth beer** -- The **Indian Pale Ale** -- (**IPA**) brewed in **India**. The **company has launched** the **beers** in Delhi-NCR, Mumbai, Bengaluru and Pune through promotion across 250 restaurants and bars, Bira 91 said in a statement. "This is a ...

bira 91: Good news, Bira lovers! Fifth beer, IPA, launched in India ...
https://m.economictimes.com/...fifth-beer-ipa-launched-in-india/.../60923549.cms?... ▾
NEW DELHI: Premium **beer** maker Bira 91 today said it **has launched its fifth beer** - The **Indian Pale Ale** - (**IPA**) brewed in **India**. The **company has launched** the **beers** in Delhi-NCR, Mumbai,

Figure 11. Answers as obtained from an Internet search

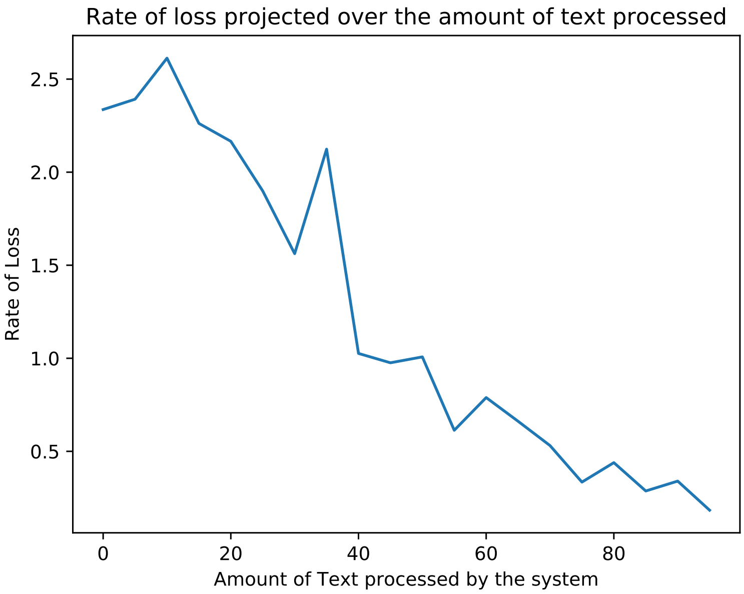


Figure 12. The Rate of Loss over text processed for the first question

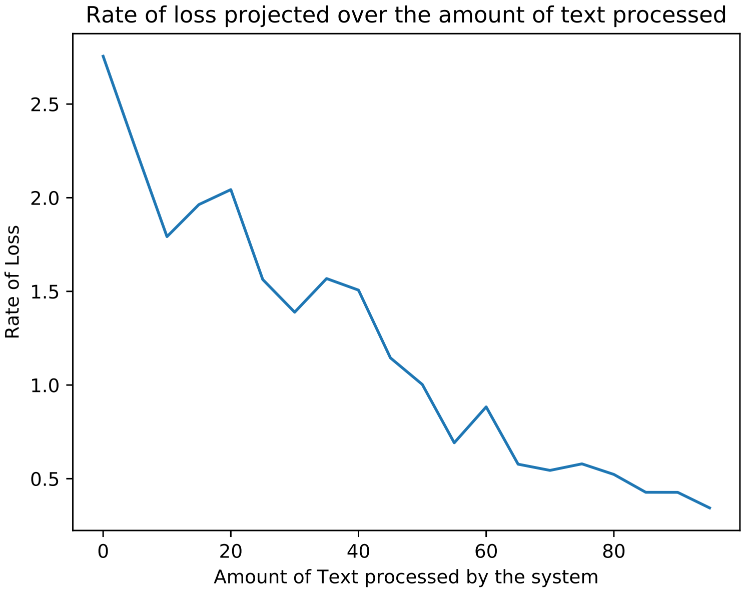


Figure 13. The Rate of Loss over text processed for the second question

Thus, in a similar manner, a total set of twenty-five questions was obtained from the user. And for each run, a specific number was chosen (for the first run, one question was chosen, on the second run, five questions were taken, and so on). Each run of the whole operation was performed for five times, per set of questions. The accuracy score was calculated as the average number of times the right answer was predicted.

Table 1. Performance of the Char-RNN for a Range of Queries

Number of Questions	Accuracy Score
1	96%
5	90%
10	80%
15	78%
20	74%
25	74%

Upon examination it can be noted that, as the number of number of questions increases, the accuracy falls, but it eventually squares off at 74%. Upon multiple runs, it was found that the accuracy score averages out to the same levels of accuracy, that is aforementioned.

The area where the accuracy of the model falls is, if the answers were to contain a numerical value embedded in it, then the model would not able to predict those accurately. An example of these type of answer keywords would be "350m", or "12ft". If the model has to predict these type of answer keywords, they need to occur enough number of times be an observable pattern for the model to pick up on, and the probability that such type of keywords occurring, is quite less. But, if they are split (such as "350" and "m"), then the numbers and the attached string parts individually have no meaning, and only add to the noise when training the model. Hence, the accuracy falls in these types of questions.

The other observation is that, as the number of questions increase, the accuracy falls. This is because there would be certain questions that would have numerical-embedded answers in them, and when cleaning or filtering them out, one would lose the vital information, which would not help the model. Thus, leading to the accuracy score dipping in value.

5. CONCLUSION AND FUTURE SCOPE

In terms of future scope, one of the main points where the accuracy of the model can be improved is in terms of the data that is being provided to it, and the training being done on it. As mentioned, the fall in the accuracy score occurs when the answer keyword contains an embedded numerical value in it. Thus, a possible future improvement for this research would include allowing the model to train on data with embedded numerical values, so that the accuracy score of the model could be improved.

On a concluding note, the authors would like to say that it was an exciting to delve into this exciting research, and was an interesting concept to explore. While performing this research, the authors learnt about various tools and concepts that are used in the research, and in the academic spheres. They were exposed to, and learnt various concepts of Machine Learning and Natural Language Processing.

REFERENCES

[1] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks," in *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 1017–1024.

[2] D. Eck, "A first look at music composition using lstm recurrent neural networks."

[3] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription," *ArXiv e-prints*, Jun. 2012.

[4] I. Sutskever, G. E. Hinton, and G. W. Taylor, "The recurrent temporal restricted boltzmann machine," in *Advances in Neural Information Processing Systems*, 2009, pp. 1601–1608.

[5] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading wikipedia to answer open-domain questions," *CoRR*, vol. abs/1704.00051, 2017. [Online]. Available: http://arxiv.org/abs/1704.00051

[6] A. Kannan, K. Kurach, S. Ravi, T. Kaufmann, A. Tomkins, B. Miklos, G. Corrado, L. Lukács, M. Ganea, P. Young, and V. Ramavajjala, "Smart reply: Automated response suggestion for email," *CoRR*, vol. abs/1606.04870, 2016. [Online]. Available: http://arxiv.org/abs/1606.04870

[7] A. Jaech and M. Ostendorf, "Improving context aware language models," *CoRR*, vol. abs/1704.06380, 2017. [Online]. Available: http://arxiv.org/abs/1704.06380

[8] S. Ghosh, O. Vinyals, B. Strope, S. Roy, T. Dean, and L. Heck, "Contextual LSTM (CLSTM) models for large scale NLP tasks," *CoRR*, vol. abs/1602.06291, 2016. [Online]. Available: http://arxiv.org/abs/1602.06291

[9] J. Tang, Y. Yang, S. Carton, M. Zhang, and Q. Mei, "Context-aware natural language generation with recurrent neural networks," *CoRR*, vol. abs/1611.09900, 2016. [Online]. Available: http://arxiv.org/abs/1611.09900

[10] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Józefowicz, and S. Bengio, "Generating sentences from a continuous space," *CoRR*, vol. abs/1511.06349, 2015. [Online]. Available: http://arxiv.org/abs/1511.06349

[11] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," *CoRR*, vol. abs/1509.00685, 2015. [Online]. Available: http://arxiv.org/abs/1509.00685

[12] A. Graves, G. Wayne, and I. Danihelka, "Neural turing machines," *CoRR*, vol. abs/1410.5401, 2014. [Online]. Available: http://arxiv.org/abs/1410.5401

[13] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *Trans. Neur. Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994. [Online]. Available: http://dx.doi.org/10.1109/72.279181

[14] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1162/neco.1997.9.8.1735

[15] M. Honnibal and M. Johnson, "An improved non-monotonic transition system for dependency parsing," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*. Lisbon, Portugal: Association for Computational Linguistics, September 2015, pp. 1373–1378. [Online]. Available: https://aclweb.org/anthology/D/D15/D15-1162

[16] E. Loper and S. Bird, "NLTK: the natural language toolkit," *CoRR*, vol. cs.CL/0205028, 2002. [Online]. Available: http://arxiv.org/abs/cs.CL/0205028

[17] S. Bird, E. Klein, and E. Loper, *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. O'Reilly Media, 2009. [Online]. Available: https://books.google.co.in/books?id=KGIbfiiP1i4C