

Transforming the Generative Pretrained Transformer Into Augmented Business Text Writer

Faisal Khalil (✉ fkhalil@uni-osnabrueck.de)

Osnabrück University

Prof. Dr. Gordon Pipa

Osnabrück University

Research Article

Keywords: Natural Language Generation, Transformers, Business text generator

Posted Date: December 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1170589/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Transforming the Generative pretrained Transformer into Augmented Business Text writer

Faisal Khalil · Prof.Dr.rer.Gordon Pipa

Received: date / Accepted: date

Abstract This study uses transformers architecture of Artificial neural networks to generate artificial business text for a given topic or theme. The implication of the study is to augment the business report writing, and general business writings process with help of Generative pretrained transformers (Generative Pretrained Transformer ([GPT](#))) networks. Main focus of study is to provide practical use case for GPTs models with help of big data. Our study model has 355 million model parameters and trained for three months on GPU enable devices using 2.3 billion text tokens(is available as open-source data now). Text tokens are collected with help of rigorous preprocessing, which includes; shortlisting of Subreddits of Fortune 500 companies and industries, listed on US-based social news aggregation online portal called “Reddit”. After shortlisting, millions of submission of users during the five years, are parsed to collect the URLs out of it. 1.8 million working URLs are scrutinized. Business text is parsed, cleaned, and converted into word embeddings out of Uniform resource Locator ([URLs](#)). The result shows that both models; conditional interactive and random sampling, generate text paragraphs that are grammatically accurate and stick to the given topic.

Keywords Natural Language Generation · Transformers · Business text generator

1 Introduction

With the passage of time, the field of artificial intelligence, and machine learning have been made progress by leaps and bounds. Nearly all fields are getting benefits

PhD candidate at Institute of Cognitive Science, Universität Osnabrück, Germany
Wachsbleiche 27, D-49090 Osnabrück
Tel.: 49-(0)541-969-3380
E-mail: fkhalil@uni-osnabrueck.de

Prof.Dr.rer.Gordon Pipa Professor and Chair of Neuroinformatic Group at Institute of Cognitive Science, Universität Osnabrück, Germany
Wachsbleiche 27, D-49090 Osnabrück
Tel.: 49-(0)541-969-3380

from the cutting-edge technologies to leverage their processes, and Deep learning is one of them. Big tech giants are reformulating their strategies to align with AI and ML. Deep learning is a branch of Machine learning that enhances the model learning process with its deep layered architecture. Like many other walks of life, Deep learning has won its spurs as a very effective and efficient technique for natural language processing related tasks. Since, computers are unable to understand the natural language, enabling them to understand the natural language and to process the information in a useful fashion has long been under the researchers' and practitioners' focus.

This study is inspired by the new method implement by the Google Brain team [Vaswani et al. \(2017\)](#) and the work of OpenAI [Radford et al. \(2019\)](#). Before introducing the transformers implement by the above-cited research work, it is important to shed the light on the recent past of Natural language processing (NLP). Although Natural language Processing (NLP) has deep roots in the past and the first breakthrough was the well-known paper of Alan Turing 'Computing Machinery and Intelligence' [Turing \(2009\)](#), real progress in the field has been made in the late 1980s - when machine learning algorithms came into the picture. The machine learning revolution has permanently changed the approaches to address NLP related problems. At the start, mostly much stress has been given to rich text features embedding - to enables Artificial Neural Networks (ANNS) to understand the rich text in numerical form. Later these embeddings are given to an end-to-end neural network that essentially maps the input and output, i.e [McClelland and Rumelhart \(1989\)](#). Later one, seminal work published related recurrent neural network [Rumelhart et al. \(1985\)](#). Recurrent models are very important for natural language processing because natural language carries lexical, syntactical, and semantic context in it- thus previous words or characters are very important to solve machine translation and text prediction tasks. In the year 2002 Jürgen Schmidhuber and his students [Gers et al. \(2002\)](#) came up with a better idea for neural network application that involves long-term dependencies, named, Long Short Term Memory (LSTM). Long Short Term Memory (LSTM) devises some gating and sates mechanism that keeps import information from the previous sequence and also memories the previous state that finally accumulates to the current state to predict the next sequence. Many enhancements have been made by the research community in the recurrent neural network model. The most highlighted models are seq2seq (sequence to sequence) [Sutskever et al. \(2014a\)](#), [Jacovi et al. \(2018\)](#). Seq2seq models essentially work with encoders and decoders recurrently to encode the output of the previous sequence and combine it with the current input. The next enhancement in recurrent model is attention mechanism, see [Xu et al. \(2015a\)](#), [Yang et al. \(2016\)](#). Attention mechanism has been proven very well in machine translations, where two pairs of sentences of two languages are mapped together with encoders and decoders.

So, looking back to the short history of the evolution of the natural language processing techniques, we understood one common limitation of all these models concerning solving the NLP task is the models are computational resources hungry and very slow. NLP corpus normally involves an enormous amount of training data, long-term dependencies, and recurrent nature. These factors make the training process very slow to achieve the desired result. Addressing this problem, the research community has come up with multilayered attention head and encoder decoders - formally called Transformers [Vaswani et al. \(2017\)](#). The current study

uses a similar approach to generate the domain specific text, and detailed methodology is discussed in 3.2. We have used a recently developed transformer neural network architecture. This architecture is primarily used for Google translation works in two different blocks, namely, encoders and decoders. We have only used the decoder part. We have provided the model with a 2.3 billion text token during the training. The model has 355 model parameters and has been trained for 3 months to reach a 2.6 training loss value. Above-mentioned 2.3 billion text tokens are collected after rigorous data preprocessing steps. US-based social news aggregation and discussion forum has been selected for data collection purpose. Almost 700 Subreddits are shortlisted for the purpose of getting URLs out of it. Millions of submissions for five years have been considered. Submission means any post, comment, or reply by the user. Users often redirect towards URLs for clarification. So, 1.8 million URLs are collected from the submissions, and validation and functionality of all URLs have been confirmed. With the help of a parser, these URLs are parsed and cleaned to get the text. Finally, 2.3 billion ready to feed to the model word embedding has been generated. In rest of the paper; literature review, Methodology of the study and model, results of the study and limitation and future suggestion have been given respectively.

1.0.1 Research gap

After getting the flashback of the evolution of the NLP and recent developments of NLP, we can see one common problem for all Natural language understanding problems is creating a relationship matrix between the words or characters and giving importance to the specific word at a specific place. Solving this problem is very important for all NLP-related niches, for example, Natural language understanding, Natural language generation and, machine translation. In this connection, we have mainly two problems to be solved. Problem no 1 is again giving importance to the words and specific place in the sentence and creating correlation or context to each word embedding based on their usage. The second very problem is supplying a lot of data or in other words a lot of instances to the model to learn the placement and relational pattern of the characters or words. Giving a lot of data needs a lot of words' embeddings matrix that leads to extremely slow model training and a lot of computation resources. So, the computational and efficiency problem is more lethal as it seems to get a breakthrough of problem No. 1. The research community either could wait for the computation resources to get more efficient and faster enough to solve the problem at hand, or they must have to come up with an optimal solution. So, the solution to this problem was attention mechanism Vaswani et al. (2017) and most specifically transformer architecture of neural networks, formally called encoders decoders Radford et al. (2019). well, fair enough transformer can, theoretically, overcome the above-mentioned problems and give a new horizon to the landscape of NLP and NLG, but we need to provide a lot of real-life use cases and proof of concept to supplement this new ANNS architecture. After this conceptual breakthrough, the next challenge is to come up with a lot of data and preprocess that much big data to supply it to these new models to proof the concept of the conceptional invention. Our paper is exactly filling this gap here by coping with the challenge of developing the proof of concept and practicality of this new advancement of NLP and deep learning. So, in this journey the most important step is to find a use case; so, we have cho-

sen business-related reports and text writing. In the next subsection, we will give precise details where and how this concept can be used in a commercial setting and what benefit it can promise. Coming back to the current point, getting a lot of business-related data is very important as well very hard because of a lot of irrelevant text and without the authenticity of being business text. So, involving humanized efforts to tag data is very costly and not plausible. So, we decided to use “reddit” a platform, widely used, and each post is voted by the community. In this way, we could get human checked data in huge volume, related to the business problems. it is also relevant to mention here that we did not parse data from “reddit” directly, rather we have only collected URL links from the posts, and then we parse complete URLs text. So, our main contribution here is rather less on the theoretical side and more on the practical side. As we have retuned and adopted the existing theoretical concept in a more practical setting to provide its proof of concept. after having this discussion, it’s very relevant to provide one hypothetical application instance and possible commercial usage of this study. So, next subsection talks about the hypothetical ideal use case and overall generic use cases of the study.

1.0.2 Hypothetical use case:

Let’s here create a practical scenario. In the office and business management, there are a lot of reports and text writing, for example, Manager X has to give a job placement ad for a consultancy firm, or, he has to write an advertisement. He has to write a small report about his product and its competitor in the industry he is operating to get external funding. In such cases the grammar is not only an important factor but are pinning words other people are using in the industry to influence more or clarity of text is maybe more important. Let say a software application helps Manager X in two ways; first, gives a context or appropriate usage of words replacement based on millions of other use cases already people used in similar instances. Second, if he writes “Apple Inc.”. the application suggests him, i.e., “Apple has launched iPhone pro max. in 2020 that gave them xxx hundred thousand \$ annual revenue”. So, now Manager X can save a lot of time and energy in surfing google in searching facts and figures. if some assistance is provided on how he can paraphrase any keywords, could improve business writing greatly. I know that requires a lot of work on front-end development too, but the Black box part would be NLG here.

1.0.3 Practical Implication:

The study has great potential for real-world practical uses: for example, next-word prediction, topic modeling to extract text out of scanned images, contextual soundness of the business writing, and suitability of word usage even if it’s grammatically correct in the first place. Any subject-specific knowledge, language usage, and vocabulary are always different compares to generic languages. Many companies and start-ups have software applications that are using a similar approach but use general language text. Here is a list of some: Gmail salutation and common words autofill used during the email [GoogleEMail \(2021\)](#), Grammarly [Grammarly \(2021\)](#) gives words context suggestion and content clarity based on the text they have trained upon. At the start of registration, they asked for purpose of use.

Maybe something like Grammarly business writer or something similar could be the very practical use of this study. Reverso Translator gives translation based on the frequency of usage of the word in literature along with text, except where the looked-up words have been used. There is the potential of usage of such tool is there where one can give the accurate context of the only business-related text. Lastly, we did know at the time of conducting this research, but one online platform emerges now which is using augmented writing approach with greater success having a top-level firm in their customers' portfolio, i.e. see [Writing \(2021\)](#). This would be a very true practical usage of such a study.

2 How deep learning integrate into corporate sector?

The literature on the Natural Language Processing is root back in the 1940s. After parsing the literature, the evolution of NLP can be segregated into different phases; for example, the journey started from machine translation problems, followed by the computers and information technology revolution - that triggered the AI applications into this area. After AI and machine learning came into the picture — complex task solving ability has been improved with less time — thus grammatical structure has been focus more. After advancements like deep learning and reinforcement learning, NLP has now entered into artificial text generation and generated text is hardly differentiates from human written.

Though the research community of that time had been working on NLP, the first scientific paper was published by the MIT language department head, William. N Locke and A.Donald Booth, head of the Brick-Beck collage [Locke and Booth \(1956\)](#). Machine Translation (Machine Translation ([MT](#))) started with three dominant languages of that time, English, Russian, and a bit of Chinese. Computational resources were too scarce and much effort had to be exerted on converting data in bits [ALPAC \(1966\)](#). Early birds in this area have given focus to syntactical computational processing of language, and it was important to first draw the basic structure for the language [Plath \(1967\)](#). Work of [Ceccato \(1967\)](#) some researchers have tried to shift the focus from the syntactical to semantic oriented language processing. Ceccato tried to co-relational analysis between the same pattern of a pair of languages and tried to achieve the semantic driven language processing. [Winograd \(1972\)](#) and [Woods \(1978\)](#) have seen the 1960s transformational grammar theory is a misfit of computational grammar and analysis and not offering much in terms of semantics. The computational confidence approach is given by Woods' and Winograd's enriched the previous work in a semantic path.

Later on, in the 80s, AI came into the picture and the community has shifted their focus toward a machine leaning based approach for solving the existing dilemmas of NLP in a pure semantics way [Schank \(1980\)](#). In this decade, researchers have realized that the NLP task such as building the word representation to use in AI-related networks and pinning the context is very hard. Some note able work of the 1980s is as follows: Briscoe et al [Briscoe et al. \(1987\)](#) have built a general-purpose grammatical formalism including syntactical analyzer for the English language with help of suboptimal software, named Grammar Development environment (Grammar Development environment ([GED](#))). They also program software to build and manage a large grammar base. Towards the direction of speech recognition, Young et al [Young and Chase \(1998\)](#) have led to major US

speech recognition projects, called, Continuous speech recognition (Continuous speech recognition (CSR)) and (Long vocabulary speech recognition (LVCSR)). The paper includes tools and methods for news transcription, text dictation, and transcriptions.

The next phase of the NLP development is the 1990s, that mostly focuses on a combination of lexical and syntactical approach for natural language processing. After lot of twists and struggle of almost two decades, the statistical and probabilistic approach has been adopted for classification tasks in NLP Sparck Jones (1992). Later on, these models became raw sources of machine learning related techniques to solve the NLP complexities. for example, Manning and Schuetze Manning and Schütze (1999) have worked on information retrieval, feature extraction out of it, and analyzing the textual information with statistical models. Mani and Maybury Maybury (1999) have used terminological logic to built a knowledge base for automatic information extraction and text summarizing. By the end of the 1990s, dialogue speech system and language processing had expanded the horizon with multilingual text machine translations, speaker-independent speech to speech dialogue system. Wahlster (2000) has worked on project Foundation of Speech-to-Speech Translation- so-called, 'Verbmobil'. This multilingual (German, English, and Japanese) takes input in a speaker-independent manner and translates them into other desired languages. it also handles domain-specific business spoken dialogues and translates into other languages with approximately 80 percent accuracy. The struggle of many years make the NLP researchers, practitioner, and industry realize that linguistic resources are inevitable for the further development in this filed, thus, two institutions, "British National Corpus" BNC (2020) and "WordNet" Fellbaum (1998) are come into being. The next era of natural language processing started after 2001. Though many models have been proposed by the researchers which were other than neural networks, we are only discussing the neural network-oriented important models in this paper.

Bengio et al Bengio et al. (2003) proposed tri-gram state-of-the-art neural probabilistic model. They have used a neural network for the probability function. The idea is based on the conjecture that unseen words get a higher probability to be predicted based on the similarity of the words - on which the network is trained. The next word prediction approach has many practical uses commercially, for example, see the work of Kannan et al. (2016) that can generate a small short semantic reply of the email.

The next advancement in the field of NLP is multitask learning, off-course this method is not only confined to the NLP but a general enhancement in the neural network world. Collobert and Weston Collobert and Weston (2008) have tried to implement this technique for transfer learning. Vector representations of the words have been fed as an input to the model to do word prediction and then learning of the current model was transferred to the other independent model to achieve a similar but not the same task. The multi-task learning approach was first introduced by the Caruana Caruana (1998). Once, so-called, word vector representations are fed to the neural network, they start learning the context and association of each work with the other. Transfer learning makes it possible to share the learned weight across the models for generalization and incremented learning approach. During the optimization process, it is very important which parameter to transfer. Ruder Ruder et al. (2019) proposed that the sharing parameter can also be learned during the learning process. See also similar research McCann et al.

(2018). In this connection, the next milestone was “vectors representation” of the text, so-called word embeddings. This basic word embedding idea was first floated by mikolov Mikolov et al. (2013). They have proposed that removing the hidden layer while training the word embedding is giving more promising outcomes. Later on, this idea paved the way for the concept ‘word2vec’ and originally adapted to two popular approaches, namely, bags-of-words and skip grams. This phenomenon has triggered the research interest in this direction and many researchers have enrich this concept see; Arora et al. (2016), Mimno and Thompson (2017), Antoniak and Mimno (2018), Wendlandt et al. (2018). The current direction of the word embedding is to train a very large corpus and use used pre-trained embeddings for multilingual models in an independent and unsupervised fashion. for example, see. Artetxe et al. (2018), Conneau et al. (2017), Søggaard et al. (2018).

In the year 2013 and 2014 neural network architectures are being applied to NLP, the most obvious choice was recurrent, recursive, and convolutional neural networks. simpler Elman Elman (1990) RNNs were replaced with LSTM by Hochreiter and Schmidhuber (1997) because of long-term context dependencies in input text. secondly, convolutional networks are originally dealt with computer vision areas but also implemented in NLP for example see the work of Blunsom and Grefenstette (????) and Kim (2014). The obvious plus of the using convolutional network is they are more parallel and local context based on layers rather than past state contrary to the LSTMs.

Concerning recurrent neural networks, the next enhancement was a sequence to sequence modeling (seq2seq). Seq2seq model is using the same recurrent architecture of the neural networks, but the important bit is disguise in encoding and decoding procedures. The input sentence is first encoded into a vector representation. The decoder then tries to decode the predicted symbols based on the encoder state sequentially. The sequence to sequence model was proposed by Sutskever et al Sutskever et al. (2014b). Later on, in the year 2016 Google Google (2020) has decided to change its monolithic sentence based machine translation to complete neural network-based. Now, seq2seq models are the foundation of language generation models and further developments, i.e transformer-based neural network architectures. Similarly, image captioning Vinyals et al. (2015) is using the same technique to generate the image captions automatically. The seq2seq model leads toward attention mechanism and transformers based approaches. The basic limitation of the seq2seq network is that it tries to compress the whole sequence of the sentence and then convert it into a fixed-length vector. Thus, the model cannot look into the hidden state. Attention mechanism, by contrast, looks into the hidden state of the model combine them to realize how much stress should be given to a specific word. Attention Bahdanau et al. (2014) was the core innovation in the field of neural machine translation that permanently replace the traditional methods of machine translation. Have a look on different flavors of attention based networks and their application; reading comprehension Hermann et al. (2015) , entity parsing Vinyals et al. (2016) , image captioning Xu et al. (2015b).

The pretrained model has gain popularity among the NLP research community. The main advantage of the pretrained model is that it is context agnostic and unsupervised model. Labeling for the NLP task can be very costlier and challenging. So, the pretrained model captures the meaning and context of one language and the leanings can be transformed into the other language to get the meaning and context generation or translation. The pretrained model was first proposed

by Dia and Le [Dai and Le \(2015\)](#). The current study is also based on pretrained multi head attention based model.

3 Methodology

In this section we have described how data is preprocessed and then processed data is fed to the model is discussed in detail. The completely preprocessed data will be available as an open-source data for further research and development.

3.1 Data Preprocessing

In this section, we have described the process of data preparation for model training. Everything else with respect to the neural network model is similar to many other applications of ANNS, but the main concept here is to leverage the training process with an enormous amount of training data. Websites could be the potential source of a lot of textual data as well as a great deal of diversity in it, but the bottleneck with websites' data is the validity of data and too much unnecessary information in it. Following the research by [Vaswani et al. \(2017\)](#) we have adopted a similar approach and choose 'Reddit' [reddit \(2021a\)](#) - a USA based social news aggregation and discussion platform with 330 million users [reddit \(2021a\)](#) to collection the website URLs to parse the data form. To ensure the validity and usefulness of the web URLs , only those links have been taken that contained more than 3 'karma'. 'Karma' is so-called assurance given by the other user about the validity of comments and discussion. In this way, we have got a human level quality check on the data. Once we have devised the mechanism of data quality, the next filer was to get the URLs that are only related to the business and Fortune 500 companies . Most of the top 500 companies have their discussion and news profile on 'Reddit' called 'Subreddit'. 'Reddit' has a very large community and thus, thousands of submissions are committed on a daily basis. The raw data, ranging from 2005 to 2017, is first programmatically collected with help of the 'Reddit' programming interface [reddit \(2021b\)](#) and stored in the 'BigQuery' database. In the next step, we have extracted all the URLs having 'karma' ranking more than 3 from the daily submission of the users. These URLs are verified, whether they are working or not and at the end 1,852,482 working URLs list was prepared to parse the textual data from 'Hyper Text Mark Language ([HTML](#))' tags. With the help of parallel computing and a computer grid, 20 GBs of text files have been collected from all working URLs. These 20 GB text files are gain filtered for some unnecessary characters and symbols. Finally, the 2,302,554,291 text token were collected to be converted into word embeddings. The process is shown in figure [1a](#) that depicts a flow of data preprocessing with help of a schematic diagram. preprocessing involves:

3.2 Methods

Next comes the transformer neural network model applied to preprocessed data. The Transformer model takes all words tokens are encoded into words embeddings,

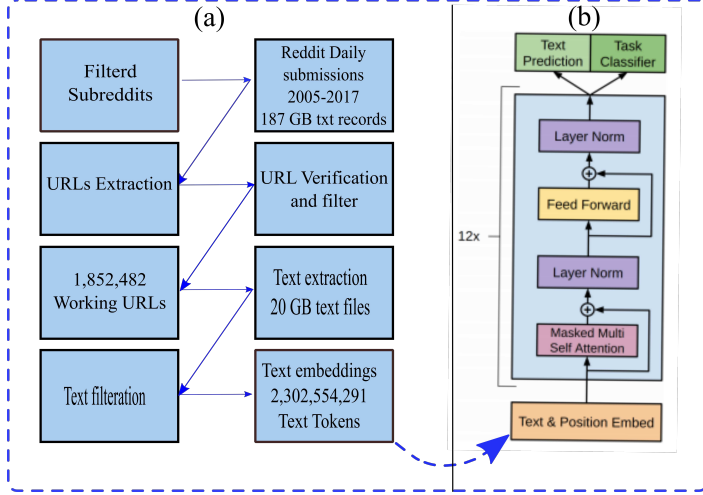


Fig. 1 Data preprocessing and network architecture

that is nothing but the numbers that represent each word. Normally, transformers have two parts, encoders and decoders, but we have only used the decoders part of the Transformer because both encoder and decoder are feasible for machine translations- that is not the case in this study. Words embedding are stored and converted into NumPy zip format for simplicity purposes. first, we will see the high-level representation of the model, and then we will look into how the self-attention layer is working. The model gets the words embedding as input, it assigns positional encoding to each word. The positional encoding keeps the position of the word into a sentence to capture the context efficiently, contrary to random order. Word embedding along with its positional information passes through the self-attention layer. The self-attention layer is twelvefold layers. For analogy purpose, we can say this layer create many copies of the sentence and map the relationship and importance of each word in the sentence to figure out how much attention to the specific words is to be given. That is why it is called a multi-head self-attention layer. We can plunge into the self-attention layer to see how it is working. Input vector $\mathbf{X}_1 \dots \mathbf{X}_N$ is multiplied by three different vectors, namely, Query vector (\mathbf{q}_1), Keys vector (\mathbf{K}_1) and value vector (\mathbf{V}_1). The vector is random weights of dimension 64 and the output of these matrices' multiplication is $\mathbf{W}^Q, \mathbf{W}^k, \mathbf{W}^v$. In the next step, we get the dot product of $(\mathbf{q}_1 \cdot \mathbf{K}_1 \dots \mathbf{K}_N)$ for sentence $(1 \dots n.)$. To stabilize the gradient process, each output is then divided to the $(\sqrt{d_k})$, whereas, d is dimension of the vector k . This operation gives us scores for each word. higher the sores means that more attention should be given to that word. In the next step all the scores for on word related to all other words should be summed up into a variable \mathbf{Z} :

$$\mathbf{Z} = \text{softmax} \left(\frac{\mathbf{Q} \times \mathbf{K}^T}{\sqrt{d_k}} \right) \times \mathbf{V} \quad (1)$$

This is the final calculation of one out of many self-attention layers, that is to be fed - in a matrix shape, to the feed-forward neural network. To focus on different

positions of the words in the sentence we need, multiple representational subspaces, subspace is achieved with the help of multiple head or copies of the attention layer. so ;

$$\begin{aligned} \mathbf{Q}_{i....}\mathbf{Q}_n &= \mathbf{W}_i\mathbf{X} \\ \mathbf{K}_{i....}\mathbf{K}_n &= \mathbf{W}_i\mathbf{X} \\ \mathbf{V}_{i....}\mathbf{V}_n &= \mathbf{W}_i\mathbf{X} \end{aligned} \tag{2}$$

whereas, $i...n$ is the number of attention layers. $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ is the query, key, and value vector and \mathbf{X} is the word embedding input matrix. So, every attention layer produces a \mathbf{Z} matrix and depending on how much attention layers being chosen, in our case 12. The attention output matrices $\mathbf{Z}_1....\mathbf{Z}_{12}$ are multiplied with the weights' matrix jointly for all layers, called \mathbf{W}_O . The resulting matrix is input for a fully connected feed-forward network. The final output of the feed-forward network is then decoded back to the words to generate the sequence of the sentence.

4 Results

In this section, we have described the results of our study. In this section, we have presented text samples that are generated by our trained model. The results include a sample from both conditional and unconditional samples. Conditional sampling means that we have provided a certain keyword to the model as an input and the model has returned a text paragraph related to that given keyword, however, unconditional means random samples generated by the trained model. Training loss summary of the 'Tensorboard' model is given in the Appendix in section 8. To support out the accuracy of model and the sample are not appear out of chance, we have given 100 randomly generated sample by the model in the appendix under section 8

We have trained the model up to 460,000 steps. Since the model has almost a 355Million model parameter and more than 2.3 billion text token, the model requires extremely excellent computation power and time. The model has been trained for 3 months on a single GPU and settles on a loss value of 2.6. This value of loss for the text-based model is quite reasonable because the language model always involves complex grammatical chains like dependencies and structures that are not easy to capture. The next two subsections provided real-time model generated text, both based on conditional and unconditional random outputs.

4.1 Interactive Conditional outputs of the Model

This subsection provides 5 different output samples of the interactive conditional sampling method of the study model. This is so-called interactive model outputs, in which the model communicates with the user. The user gives input/keywords to the model and the model generates a text paragraph that mostly talks about the given keyword/topic. Given below are the table boxes that show output against five different user given inputs.

Table 1 Results of Interactive conditional samples

Key:	conditional output:
Apple iPhone	7 Plus: 3-month free trial 15% off \$150+ purchase iPhone 7 Plus: 3-month free trial 15% off \$150+ purchase 4G LTE: iPhone 6s: 2 year data plan 3 GB data, \$49.99/mo 8GB data, \$59.99/mo 4G LTE: iPhone 6s: 2 year data plan 3 GB data, \$49.99/mo 8GB data, \$59.99/mo iPhone 5s: 2 year data plan 5 GB data, \$45.99/mo 8GB data, \$69.99/mo

output generated by the trained model against user given topic/key.

Table 2 Results of Interactive conditional samples

Key:	conditional output:
oil&price	“which make up the global oil industry. The used state-of-the-art computer models to show that the price of crude from North Dakota has increased by a factor of 40 over the past year. In the past year, it has risen by 28 percent, a period of unprecedented energy pricing, according to the report. That is an almost unheard-of percentage increase, said Tim Leissner, an assistant professor at the University of Michigan who specializes in energy and economic modeling, which is used in energy forecasts. “Usually, the percentage increases are pretty much the same year after year, so that’s a pretty consistent pattern,” said Leissner, who was not involved in the report. “Normally what we would expect is that the price goes up because of oil production, and it generally goes up at the end of the year.” But that is no longer the case, said Leissner. “The oil industry has a new president in Trump, and his agenda is very aggressive on a range of issues. And the industry is not only seeing this increase, but it is actually seeing it drop,” Leissner said. The increase in North Dakota oil production has gone from 6.5 million barrels per day in the first half of 2017 to 7.3 million in the second half of 2017 and 7.3 million in the first quarter of 2018, according to the report. And in the last two months alone, North Dakota’s production has fallen from 4.6 million barrels per day in the first half of 2018 to 3.3 million in the second half of 2018. The increase in oil production in North Dakota has also affected a number of other energy prices, with natural gas prices rising about 12 percent in the first half of 2017 and natural-gas prices rising about 13 percent in the first half of 2018. The data comes as the Trump administration prepares to review the Obama rule that requires oilfields on federal lands to submit to federally supported research and exploration programs. The rule will affect oil and gas development in the state. Leissner said the oil market will be monitored for “the next few years” for the first time in decades. “This kind of increases is a big deal, given that oil is the only major product in the world that people are going to be producing for years to come,” he said. The research is published in the Journal of Energy Technology. The report, which included”

output generated by the trained model against user given topic/key.

4.2 Unconditional outputs of the model

In the section below we have given some table boxes that contain the random sample output generated by the model. this is an artificial text written by the model. If we observe the generated paragraphs, it is very clear that the text is following the grammatical rule mostly and topics of the sample pointing towards the business-related text. An enormous amount of sample can be produced on demand, due to the brevity of this article we have only given some sample.

Table 3 Results of Interactive conditional samples

Key:	conditional output:
Microsoft Windows	<p>“users, if you don’t already have them or you’re thinking about installing them on your Windows Vista/Windows 7 PC, you’ve got to pay the price. There’s also the issue that you’re going to have to run Microsoft on your current PC, which isn’t ideal if you’re a Mac user. If you only do this once, and you’ve got Mac OS X installed, Windows 7 will work. The problem with this is, it’s a pain to use on those platforms. Even if you’ve got Mac OS X on your computer, it’s going to be a pain to use in Windows, and if you want to use Windows Vista on a Mac, it’s going to be a pain on Windows. So we’re going to need to figure out a way to get Windows 7 running on existing Macs. We’ve been trying to do so, so far. And at some point, we’ve found something that’s going to let us do it. We’ve been working with the folks over at Xiph.ai and are able to install and upgrade and upgrade, and I don’t want to go into too much detail, but we actually managed to install Windows 7 on Mac OS X through the Xfce application. And if you do this just once and you’re running Windows 7 on a Mac, it will run. Xfce and Window 7 To get Xfce running for Windows, right-click in the Xfce launcher, and then choose “Xfce Installer”. This will open a pop-up window in which you can choose what to install. The first option is “Mac OS X 10.4.5”. You can choose the installer from the list below. If you’re using the installer option, you’ll notice that Xfce has been created for you. And you’ll also see an “Xfce.app” folder for Windows. In order to get Xfce running on Windows, we’ll need to run the installer from that folder. Open Xfce by going to its application menu. Select “Install Xfce” from the search box. You’ll notice it’s not”</p>

output generated by the trained model against user given topic/key.

Table 4 Results of Interactive conditional samples

Key:	conditional output:
S&P 500	<p>“index fell 1% and the S&P 500 and Dow Jones Industrial Average gained 0.7% and 0.4%.The Dow closed up 8.2%.Meanwhile, the S&P 400 surged 3.8%, with the Nasdaq Composite index up 8.3%. The Nasdaq rose 0.4%, while the S&P 500 3.6%.While Wall Street was off the back foot, on the financial markets, the Dow rose 1.8% and the Nasdaq rose 0.8%.In the energy market, the Standard & Poor’s 500 index dipped 1.2%, the Nasdaq index rose 3.6% and the S&P lost 4.1%. In real estate, the S&P 500 rose 0.8% and the Nasdaq climbed 1.2%.Meanwhile, the Nasdaq’s gains were smaller in the oil energy group, while it gained 0.4% and the S&P 500 rose 0.3% in residential and business.The Nasdaq rose 2.0% and the S&P 500 gained 0.8% in the energy sector.At the top of the industry, the S&P 500 closed 6.9%, with the Nasdaq up 2.6% and the S&P 500 up 2.1%.In the energy group, the Dow rose 8.7%, with the Nasdaq rising 2.4% and theS&P500 adding 4.5% in the energy segment.And in residential, the S&P 500 rose 1.7% and the Nasdaq rose 1.3%.In the business sector, the S&P 500 rose 2.5% and the Nasdaq climbed 1.1%.The Dow was up about 7.8% as the S&P 500 fell 1.4%, while the Nasdaq declined 1.2%.The Dow hit a record high of 11,868. The S&P 500 was up 11.4% and the Nasdaq rose 0.4%.The financial sector was trading at a record high as investors worried about the economy and a potential Federal Reserve interest rate hike, which has taken hold following a sharp decline in the central bank’s benchmark interest rate.The Dow closed up 8.7%, the S&P 500 rose 3.8%, and the Nasdaq rose 3.9% ”</p>

output generated by the trained model against user given topic/key.

Table 5 Results of Interactive conditional samples

Key:	conditional output:
health care industry	<p>“in a manner that protects the privacy of patients. We will continue to take this necessary step to ensure that our industry is transparent and accountable,” the department said in its statement. In addition, the department plans to “deregulate the sale of any health care services without a patient consent,” and “work towards an industry-wide ban on involuntary sex trafficking. The Department of Justice has been involved throughout the issue of sex workers and the use of technology, and is actively working with regulators and private-sector partners to better address the issue of human trafficking, the company stated. “With recent advancements in technology for technology’s sake, sexual exploitation is no longer confined to the virtual realm, but now has become a reality,” it added. The department said it “will continue to work with state, federal and private sector partners to improve their policies and laws regarding human trafficking, including protections and the use of technology to protect victims of sex trafficking.” The government’s goal is to help ensure that “we do not, in any way, further endanger any of our citizens, particularly those whom we have entrusted with our law enforcement and public safety responsibilities.” The department added, however, that it “will not tolerate law enforcement personnel, employers, contractors and contractors to be complicit or assist in unlawful human trafficking, as that might undermine our mission.”—endoftext—“This domain is no longer available, possibly due to a change in the registrar or the domain administrator.”—endoftext—“What started as a series of tests to see if it could survive the cold, the hot and the humidity is turning into a real life experiment that will test the idea that this tiny space heater might have a practical purpose. For this project, the team set out to conduct the tests using just a standard fridge - just a standard fridge, for short. But they wanted to see what it could do if the temperature inside it dropped by 10C - a scenario that will be a reality in 2017. The first test will bring in an air temperature of -30C - something which will kill any kind of bacteria as well as fungi in about two days. As you can imagine, that wasn’t the worst of the worst in the test. A typical fridge - even the kind you’ll find in the kitchen - was just as bad as the prototype at catching bugs. The team has now developed a small freezer for the freezer, which can withstand even lower”</p>

output generated by the trained model against user given topic/key.

Table 6 Results of Non-conditional samples

Sample	Machine Generated Text
Sample1:Topic Microsoft	<p>SA-4, but some of the other models that it supports can be set, as can the version of Windows. However, most users will need to install the free and limited OS X Lion operating system. To get it the free version, you can purchase it from the Apple site. In the past few years the Windows phone market has grown significantly. Microsoft has been aggressively supporting Microsoft Office applications in Windows 8. The company is still selling a wide selection of Office applications. The reason Microsoft has decided to target the Android market, especially is that, as a company that is heavily invested in smartphone users, Microsoft will be less apt to change direction. Still, it is unlikely that you're going to be surprised to find that Windows Phone has managed to outcompete Android in terms of support level and functionality. That has been the case for years now, and not just because of the various operating systems. However, things have changed for the better. The development time has gotten slower, and the hardware has gotten more modern. There have also been increased efforts to make the operating system itself more user friendly. Microsoft has been steadily working to increase the range of features available on the platform. In the case of Windows Phone 8, this means that it supports the latest version of the operating system, the Universal Windows Platform (UWP), which has evolved in ways that made the system more accessible to the new users. The new version of the Windows Phone OS also integrates with the new Metro UI, which has been available for several years already. This means that the interface is easier to navigate with each new update. There is an obvious difference between the Windows Phone version of the Windows OS and other systems, which makes the difference in the Windows Phone OS much stronger. As a result, the Windows Phone OS is likely to enjoy a much wider appeal. The Windows Phone OS is a much more mature OS, however, and it may prove to be even more attractive once it is officially supported by Microsoft. In that case, it seems that the Windows Phone OS can only prosper as long as Redmond will provide more hardware devices that can be run this OS. It is, of course, a very hard problem to solve. However, it seems inevitable that this issue will play a greater role in Microsoft's future strategy. Microsoft has been focused on offering a wide range of popular consumer and enterprise computing options in order to take advantage of the growing mobile market. There is also a good chance that the introduction of Windows and Office to the marketplace will bring a greater opportunity for Windows to become a mainstay for mobile devices. Further Reading—endoftext—A new video showing a drone carrying a baby to her birth, while also showing her running and jumping, is set to debut at the London premiere of David Cronenberg's "Puff Daddy," at the V&A on Wednesday. Puff Daddy follows a teenage girl whose father is killed in an accident and has been left with an orphaned daughter. Her ex is a young woman from a nearby suburb who has a passion for flying and is looking for a way to give back to her community. "You can't have a child without a parents," said Cronenberg, who directed "All the Money in the World" and "O Brother, Where Art Thou?" alongside Brian De Palma and John Frankenheimer, as well as "Shakespeare's Son," "All the Money in the World" and "The Other People's" alongside Tom McCarthy. The director also showed off new CGI footage of the film's main characters, including the first scene where they're shown playing with the baby and flying. "Puff Daddy" is shot in a sequence that features a close-up of the girl and the baby. "We wanted to create a visual effect in ways that were visually appealing," Cronenberg said. The scene with the daughter flying was filmed in the streets of North London, but Cronenberg said the scene in the film's last shooting, "Rudolph the Red-Nosed Reindeer," will also appear in the visual effects package. The project also features a "flying baby" sequence that was filmed in a nearby suburb.</p>

output generated by the trained model

Table 7 Results of Non-conditional samples

Sample	Machine Generated Text
Sample2:Topic: Health Care	<p>,” said the chief executive officer of the British medical charity, Beaumont Hospital. ”In recent years we have seen a surge in the number of young people coming into this country seeking to change our society, but the risks of those who do become radicalised remain too high. ”Young people like Mika are at risk of radicalisation and may be vulnerable to becoming radicalised themselves through viewing social media as a possible means of radicalising themselves or others. ”Our advice is to work closely with the police and other relevant authorities to help these young individuals to understand the risks in the communities they may come into contact with in the future and to talk to parents about their responsibilities.” It has long been feared that social media is inextricably linked to radicalisation. Earlier this year it emerged that the police were monitoring 4 million posts on Twitter, Facebook, Kik and Line, all forms of instant communication, for signs of terrorism. But Dr John Ralston of the University of Oxford has claimed that although “social media has been used for a long time in the UK, and indeed throughout Europe, some parts of society have never noticed it.” He said that while people in certain sections of the community have been concerned at the recent rise in extremism, many young people in other parts of the population have not. ”The vast majority of young people at one time or another have encountered such people through social media.”—endoftext—;The biggest financial institutions have the greatest exposure to the market, yet they are the most transparent, according to Transparency Market Research (TMR). The research group analyzed more than 700 leading financial companies, looking for those who reported some form of transparency, or disclosed more information than allowed. The findings, based on the organization’s annual survey of 1,250 U.S. companies, show that firms with the largest exposure to the financial market have disclosed the highest amounts of transparency – even though they are less transparent than the average firm. Those firms with the most transparency are also the companies that the study found to the highest risk of market abuse, including: • A number of the top 50 firms made disclosures in excess of 30 percent of their company size. • The majority of firms did not disclose their disclosure forms during the year.</p>

output generated by the trained model.

Table 8 Results of Non-conditional samples

Sample	Machine Generated Text
Sample3:Topic Energy Market	<p>to have more time with people,” he says. “I am sure I will enjoy sitting on my porch with the trees behind me. It will be a really relaxing time.” For all the progress the film has made this year, however, as the box office has surged, so has that of the franchise. The latest installment, starring Tom Hiddleston, will bow on Monday, while the two-part drama “Hiddleston: Longmire” will arrive in U.S. cinemas on February 19, followed by its worldwide debut in April. —endoftext—;Hodl has been selling the technology for years. On the shelf of any Walmart, Walmart.com or Amazon, it’s not uncommon to come across a shelf full. But now a new technology to turn your clothes into a new energy-generating asset comes to San Francisco and Silicon Valley. As the world warms and urban temperatures rise, the amount of energy stored in the fabric of the clothes and other products increases. This process has the potential to revolutionize a whole range of industries and technologies—as well as the way people buy and use clothes. Firms that design, produce, market, sell or install this technology face a number of problems, including the technology’s limitations—which includes the inability to make use of the wind. And in a major market, technology can be confusing to customers. If an installer sees a new technology on a product, it’s not clear where to go. Is it a product or an energy-generating product? It’s an industry with a number of big names, all of which are focused on the same thing, but each company has a different brand. Sprinting out of the closet Fifty years ago, it was a relatively new idea that was made possible at the beginning of the Cold War by advances in the nuclear power and hydrogen bomb. This meant that we could get more power with less fuel than any previous technology had before. But more power, as it turns out, is less efficient than it used to be, and that has led to more problems than it is solving. The technology was called the “energy storage.” The technology used energy (or energy storage) to cool a part of fabric. It was the first technology for storing energy. Then this technology fell out of fashion. In the mid-1970s, it was popularized by Motorola, a firm that produced smart refrigerators that could store energy and heat the refrigerator. This technology became the fuel cell and energy storage industry before it fell out of favor. Sprinting out of the closet The reason we still have to solve problems with energy storage is partly due to the sheer size of the industry. It takes lots of expertise to be able to make use of the technology, but we still find ourselves getting it wrong. The biggest reason are twofold. First, the technology was extremely effective early on, but as technology has made more efficient, then it has gotten difficult to get it right. Second, the technology is often confusing and difficult to integrate into existing products. If you’re in an industry where the problem with storing energy is a lack of innovation, you may find the technology confusing and cumbersome and you may have to learn how to use multiple energy storage products to get the right kind of energy. The companies behind the two companies—Gigasolar, in Palo Alto, and GigaSolar, in Sunnyvale—are trying out new energy storage technologies that make an energy store. Sprinting out of the closet is a common solution to those issues. However, there are companies that are able to solve the first problem using their previous technology without ever looking at the second problem. One company that has done this is Gigasolar, in Sunnyvale and Stanford. Gigasolar’s company mission is to get rid of the technology gap between what the customers need and what the companies are built for. It’s not a simple business. They have to design a product that solves the problem before it can be fully put to use. The company had to do this. Gigasolar, which is run by the founders Peter and Adam Zweig, are very involved in figuring out how to work within the energy delivery space. They have a number of patents that cover multiple products. In these patents, they have some of the best energy energy storage patents that you can find. For instance: They use carbon dioxide to store energy in the atmosphere and then they release carbon dioxide back into the atmosphere to convert it into new energy.</p>

output generated by the trained mode.

Table 9 Results of Non-conditional samples

Sample	Machine Generated Text
Sample4:Topic Retailing	<p>production are not the same as production at a retail store like you are likely to see at the Superstore. I recently did a bit of testing for this. You Can't Buy This There are two types of products that can make it through the Superstore checkout process. The first are the items that are sold using the same pricing structure as the retail store. For example, the price for the T-Shirt above only goes up \$0.75 with the same color. All I had to do was go to the Superstore and double check my shirt price with the same information as I use to determine if my T-Shirt cost me half what it cost on the same item in the Retail Store. The T-Shirt above, when bought online, has been \$0.75 up charge, but on sale is \$0.85 for the same size of shirt (you can see why the \$0.75 difference is so important!). The second type of product that can help you make it through the checkout process are items with the same price that you'll find on the retail store. This is usually with items like apparel. For example, the same shirt is \$.25 on the Superstore (a T-shirt cost me \$0.85 on the Retail Store for the same size of Shirt), as long as they look the same they're actually a bargain. If you're a Clothing shopper and shop your clothing online, you'd probably get the price of those same garments for \$0.75. In my opinion, the Superstore prices for the same items are more consistent, so I've decided to include them. To find the Superstore price for Clothing, you need to purchase a Clothing App item online, take the price as your price, then click the purchase button. You should then follow the steps above to determine the price for that item. A good starting point would be the same price that you pay for the Clothing App item in the Retail Store. Here's an example of the Clothing App item that I would purchase on the Retail Store. The Shirt above, I bought for \$27 online. If you purchase this on the Retail Store, \$27 would be the price you'd pay online: \$37.99. If you'd purchased it on the Superstore it would be \$0.80. I would take my clothing price and multiply it by the retail value of the Shirt (i.e. \$0.80 is what I think the T-shirt is on sale at the Superstore). I would assume this will be the same price as my T-Shirt. To do this, I just have to multiply the actual price by my price, then subtract my retail value from my Retail Store price to create a final price of \$0.40 for the T-shirt. As you can see, my total Clothing App price is \$38.50. To keep track When you're shopping online, take in one value from each App item for those two different sizes of the Shirt (the same ones) then calculate that value and convert to a per-item price for the product you're looking at while shopping. The final price should be that. So let's say that we have the Shirt below for \$25 that we think would be the same \$25 price for the same Shirt on the Retail Store. After we calculated the per-item value for the Shirt, we would add \$0.80 to our total clothing price for that T-Shirt to calculate our final price. In this case, that final price would be \$0.80. This is the final value of that T-Shirt - \$0.80.</p>

output generated by the trained model.

4.3 Discussion

In this section, we are going to discuss the results of the study and how these results stratify the problem inference of the study. The main focus of this study is to testify the validity and useability of current theoretical development in the field of natural language generation and generally Natural language processing. As we stressed out previously the long dependencies chain of one word to other words, placement of the word in a given sentence and relational space of the word and characters is the big challenge of language generation-related problems. This problem was very difficult for recurrent neural network models to cope up with. So, the researchers came up with different theoretical concepts. In this connection, we are providing practicality, useability, and proof of concept of the model in our study. For this purpose, we have provided two types of results, interactive

conditional and non-interactive random samples. How we have trained the model, iteration, and loss graph can be observed in the appendix section. However, the main objective of the study was to generate the text that sticks to the overall topic of text, formally called topic modeling, secondly grammatically correctness, and thirdly, somehow related to business only. If we closely read the results of section 1, we have provided the model with random business-related diverse keywords from all different business genres. Model is not only able to collect very related text, but also supply some facts and figures. Moreover, the linkage of sentences and story making is very decent. That does the perfect job for the hypothetical use case and highlighted research gap. Additionally, for the robustness of the model, we have also created a random sample of text generations with thousands of instances. Due to brevity, we have provided some samples here, and we have given link to could to access all other thousands of samples. In both types, an interactive and non-interactive, model is achieving the initial goal for context, relatedness, and topic modeling. of course, this is just a founding block to generate any meaningful commercial soft application. we need to assemble other pieces of puzzles, namely front-end development, scripting, and mapping for words PACs of words matching the counts and statistics concerning the whole database, with many more bumps and stager on the journey down this road.

5 Conclusion

The current study is focusing on the application of Natural language processing in the field of business writing. In the recent past, the Deep Learning research community has come up with a new architectural style of deep layered AI models that are aligned with the specific need of natural language and text generation. The transformer is one of those models that are proven to very accurate and effective in context and grammar capturing in the text.

Response to the possible question, what is the purpose of the study very briefly? The study uses a generative pretrained neural network model. The model is fed with a lot of business-related preprocessed text data acquired parsing the 1.8 million URLs collected from Reddit. As a result of the trained model, user can give keywords or some topic to the model and model produced paragraph that completely sticks to the given topic, provided that the topic or keyword is in the domain of business or management sciences. These features or results provided by the study can be utilized in automatic paragraph prediction to assist the business report writer or any relevant person involves in the writing process. As there are many applications available for next word prediction generally but paragraph prediction is lacking.

Now, let us give little more details on how data is preprocessed, and the model is trained to get the output? A large amount of quality data is very important for language processing models. To address the quality issue we have chosen 'Reddit'; news aggregation and content sharing platform. Although Reddit covers a lot of different topics, we have shortlisted 'subreddits' - topic-specific Reddits. There is a huge amount of Reddit submission every day and 'KARMA' vote is given to the post that is helpful for the community. So we have collected 1.8 million URLs from those submissions that have 'KARMA' vote greater than three. In the next separate step, we have collected and cleaned all the text available in the URLs.

In the end, 2.3 billion text tokens have been fed to the model. The model has 355 million parameters. After three months of model training, the model can generate grammatically correct and aligned with business topic text as a model output. In the coming subsection, we have discussed what could be the practical application of the model and future suggestions along with some limitations of the study.

5.1 Implications and future work suggestion

There are many possible implications of this study. One possible use is market intelligence report writing. Possibly a piece of software can be developed to auto-complete the paragraphs for business intelligence report writing. Any business-related industry can be benefited with help of paragraphs prediction instead of just word prediction. In this way, the speed of efficiency of the user can be enhanced significantly. As for future suggestions are concerned, we think that text token prefixed by the theme or topic of the text can make this model more useful. For example during the training text, at the start of the text, we can provide what this piece of text is talking about. In this way, we can have greater control over the output of the model we can generate real-time long reports based on specific keywords. The report is just one example we can utilize the model is much more effective ways. We hope that the research community is maybe already doing something in that direction.

5.2 Limitations of the study

We have tried to do the study at our best, but there certain technical limitations of the study. Since the models related to text generation is usually based on an enormous amount of training data; that is a very important factor to capture the grammatical structure and relatedness of the topic, this study only relies upon the text generated from those web URLs that were discussed in business-related subreddits. The study may be improved significantly with help of having more sources of training data and more computational power.

6 List of abbreviations

NLP Natural language Processing
GPT Generative Pretrained Transformer
URLs Uniform resource Locator
ANNS Artificial Neural Networks
LSTM Long Short Term Memory
MT Machine Translation
GED Grammar Development environment
CSR Continuous speech recognition
LVCSR Long vocabulary speech recognition
HTML Hyper Text Mark Language

7 Declarations

7.1 Ethics approval and consent to participate

Not applicable

7.2 Consent for publication

Not applicable

7.3 Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

7.4 Competing interests

The authors declare that they have no competing interests

7.5 Funding

Not Applicable

7.6 Authors' contributions

The research study is written by Faisal Khalil during the PhD at University Osnabrueck, Germany. Prof.Dr.Gordon Pipa has supervised and improve the study in many aspects.

7.7 Acknowledgements

No Acknowledgements

7.8 Authors' information (optional)

Mr. Faisal khalil is doing PhD at university Osnabrueck, Department of cognitive sciences. His major areas are artificial intelligence and Deep learning. His main focus is application of AI in the areas of business and corporate finance.

References

- ALPAC (1966) Language and machines computers in translation and linguistics
- Antoniak M, Mimno D (2018) Evaluating the stability of embedding-based word similarities. *Transactions of the Association for Computational Linguistics* 6:107–119
- Arora S, Li Y, Liang Y, Ma T, Risteski A (2016) A latent variable model approach to pmi-based word embeddings. *Transactions of the Association for Computational Linguistics* 4:385–399
- Artetxe M, Labaka G, Agirre E (2018) A robust self-learning method for fully unsupervised cross-lingual mappings of word embeddings. *arXiv preprint arXiv:180506297*
- Bahdanau D, Cho K, Bengio Y (2014) Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:14090473*
- Bengio Y, Ducharme R, Vincent P, Jauvin C (2003) A neural probabilistic language model. *Journal of machine learning research* 3(Feb):1137–1155
- Blunsom P, Grefenstette E (????) Nal kalchbrenner, et al. 2014. a convolutional neural network for modelling sentences. In: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*
- BNC (2020) British national corpus. URL <https://www.english-corpora.org/bnc/>, accessed on : 2020-07-04
- Briscoe T, Grover C, Boguraev B, Carroll JA (1987) A formalism and environment for the development of a large grammar of english. In: *IJCAI, Citeseer*, vol 87, pp 703–708
- Caruana R (1998) Multitask learning. *autonomous agents and multi-agent systems*
- Ceccato S (1967) Correlational analysis and mechanical translation
- Collobert R, Weston J (2008) A unified architecture for natural language processing: Deep neural networks with multitask learning. In: *Proceedings of the 25th international conference on Machine learning*, pp 160–167
- Conneau A, Lample G, Ranzato M, Denoyer L, Jégou H (2017) Word translation without parallel data. *arXiv preprint arXiv:171004087*
- Dai AM, Le QV (2015) Semi-supervised sequence learning. In: *Advances in neural information processing systems*, pp 3079–3087
- Elman JL (1990) Finding structure in time. *Cognitive science* 14(2):179–211
- Fellbaum C (1998) Towards a representation of idioms in wordnet. In: *Usage of WordNet in Natural Language Processing Systems*
- Gers FA, Schraudolph NN, Schmidhuber J (2002) Learning precise timing with lstm recurrent networks. *Journal of machine learning research* 3(Aug):115–143
- Google (2020) Alphabet inc. URL <https://www.google.com>, accessed on : 2020-07-04
- GoogleEMail (2021) Gmail. URL <https://www.google.mail.com/>, accessed on : 2021-11-15
- Grammarly i (2021) Grammarly. URL <https://app.grammarly.com/>, accessed on : 2021-11-15
- Hermann KM, Kocisky T, Grefenstette E, Espeholt L, Kay W, Suleyman M, Blunsom P (2015) Teaching machines to read and comprehend. In: *Advances in neural information processing systems*, pp 1693–1701

- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural computation* 9(8):1735–1780
- Jacovi A, Shalom OS, Goldberg Y (2018) Understanding convolutional neural networks for text classification. *arXiv preprint arXiv:180908037*
- Kannan A, Kurach K, Ravi S, Kaufmann T, Tomkins A, Miklos B, Corrado G, Lukacs L, Ganea M, Young P, et al. (2016) Smart reply: Automated response suggestion for email. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 955–964
- Kim Y (2014) Convolutional neural networks for sentence classification. *arXiv preprint arXiv:14085882*
- Locke WN, Booth AD (1956) Machine translation of languages. *American Documentation (pre-1986)* 7(2):135
- Manning CD, Schütze H (1999) *Foundations of statistical language processing*
- Maybury M (1999) *Advances in automatic text summarization*. MIT press
- McCann B, Keskar NS, Xiong C, Socher R (2018) The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:180608730*
- McClelland JL, Rumelhart DE (1989) *Explorations in parallel distributed processing: A handbook of models, programs, and exercises*. MIT press
- Mikolov T, Sutskever I, Chen K, Corrado GS, Dean J (2013) Distributed representations of words and phrases and their compositionality. In: *Advances in neural information processing systems*, pp 3111–3119
- Mimno D, Thompson L (2017) The strange geometry of skip-gram with negative sampling. In: *Empirical Methods in Natural Language Processing*
- Plath W (1967) *Multiple path analysis and automatic translation*. Amsterdam, The Netherlands: North-Holland
- Radford A, Wu J, Amodei D, Amodei D, Clark J, Brundage M, Sutskever I (2019) Better language models and their implications. *OpenAI Blog* <https://openai.com/blog/better-language-models>
- reddit (2021a) Reddit. URL <https://www.reddit.com/>, accessed on : 2020-07-15
- reddit (2021b) reddit. URL <https://www.reddit.com/dev/api/>, accessed on : 2020-07-15
- Ruder S, Bingel J, Augenstein I, Søgaard A (2019) Latent multi-task architecture learning. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol 33, pp 4822–4829
- Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. *Tech. rep.*, California Univ San Diego La Jolla Inst for Cognitive Science
- Schank RC (1980) Language and memory. *Cognitive science* 4(3):243–284
- Søgaard A, Ruder S, Vulić I (2018) On the limitations of unsupervised bilingual dictionary induction. *arXiv preprint arXiv:180503620*
- Sparck Jones K (1992) *Thesaurus*. *Encyclopedia of artificial intelligence* 2:1605–13
- Sutskever I, Vinyals O, Le QV (2014a) Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*, pp 3104–3112
- Sutskever I, Vinyals O, Le QV (2014b) Sequence to sequence learning with neural networks. In: *Advances in neural information processing systems*, pp 3104–3112
- Tensorbaord (2020) Google tensorboard. URL <https://www.tensorflow.org/tensorboard>, accessed on : 2020-10-15
- Turing AM (2009) Computing machinery and intelligence. In: *Parsing the turing test*, Springer, pp 23–65

- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems, pp 5998–6008
- Vinyals O, Kaiser L, Koo T, Petrov S, Sutskever I, Hinton G (2015) Grammar as a foreign language. In: Advances in neural information processing systems, pp 2773–2781
- Vinyals O, Blundell C, Lillicrap T, Wierstra D, et al. (2016) Matching networks for one shot learning. In: Advances in neural information processing systems, pp 3630–3638
- Wahlster W (2000) Mobile speech-to-speech translation of spontaneous dialogs: An overview of the final verbmobil system. In: Verbmobil: Foundations of speech-to-speech translation, Springer, pp 3–21
- Wendlandt L, Kummerfeld JK, Mihalcea R (2018) Factors influencing the surprising instability of word embeddings. arXiv preprint arXiv:180409692
- Winograd T (1972) Understanding natural language. Cognitive psychology 3(1):1–191
- Woods WA (1978) Semantics and quantification in natural language question answering. In: Advances in computers, vol 17, Elsevier, pp 1–87
- Writing TA (2021) Textio augmented writing. URL <https://textio.com/>, accessed on : 2021-11-15
- Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015a) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
- Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015b) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
- Yang Z, Yang D, Dyer C, He X, Smola A, Hovy E (2016) Hierarchical attention networks for document classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pp 1480–1489
- Young SJ, Chase LL (1998) Speech recognition evaluation: a review of the us csr and lvcsr programmes. Computer Speech & Language 12(4):263–279

8 Appendix

The figure 2 shows a model training summary in graphical shape. The graph is produced with help of “Tensorflow” tool called “Tensorboard”. Tensorboard is a tool for viewing the hidden layers and mechanism of the ANN models - written by Google to increase the efficiency of “Tensorflow” library [Tensorbaord \(2020\)](#).

8.1 Random Samples

In the section, we have given the 'Microsoft 'OneDrive' shared folder link which contains 2,284 samples that are generated by the model during the training process. The random sample has been generated roughly after every 200 training steps. Samples can be accessed via following link:

[Click here to see the all random samples!!](#)

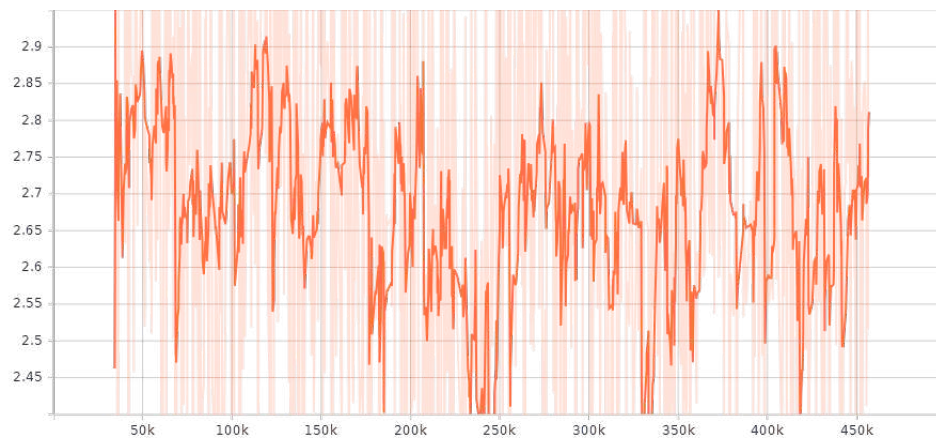


Fig. 2 Tensor-board Training loss summary