Artificial Intelligence- Modeling human like Speech and Text summaries   
CPSC-57200

Artificial Intelligence 2

Lewis University

Jacob Reed  
 jreed5@lewisu.edu

Chris Pamieta  
cpamieta@gmail.com

***Abstract— For many, the use of Artificial Intelligence is now encompassing their everyday lives without many even knowing how it works, why it works or how Artificial Intelligence is furthering their everyday way of life. The challenge with AI development is the implementation of a platform that is unique to many but useful for all. This paper focuses on what leading technology companies such as Google are trying to achieve when implementing human like voice and text recognition on their devices.   
Keywords— Artificial Intelligence, Neural Networks, Voice Recognition, Text Recognition, Summarization,Text Compression***

# Introduction

In today's day of age, when people talk about Artificial Intelligence, they think of the Roomba robot, Google Home, and smartphones. It's becoming the normal to have AI in every household, and not seen as a novelty item. Most do not fully understand the capabilities of AI, or how AI is advancing, enriching and furthering their lives. The key to having AI adopted by the masses is how well it works with everyone, and how useful it is. You might consider this the “interoperability” of AI with multiple platforms or applications. With everyone having a unique voice and look, the need for an AI application to have the capability to identify each person and be able to understand every accent, dialect, pitch, tone and a multitude of other components is key. With that, voice/text recognition, along with object, facial recognition, and human like interaction all play a very important role. Next is the usefulness of AI for everyday consumer use. As it stands, devices like Google Home play more of a passive role, and only acts in a literal way when a user interacts with. With that, the group will be looking improving human like interaction where text summarization will be the focus. As of now, two major flaws are seen in text summarization which are how informative and grammatically correct the summary is. In this paper we look to tackle the issue with grammar.

# Background

Voice and text recognition is new in a technological sense due to its vast amount of capabilities that it is applied towards. However, voice recognition and inventions that understood the voice have been around for far longer. During the late 19th Century, Thomas Edison developed dictation machines which could record user’s speech and were used greatly among doctors and secretaries. [5] The original telephone worked off similar principles of understanding not the voice but rather the sound wave patterns coming from ones spoken voice. This recognition had to be sent as soundwave transmissions through lines over distances to another user. Through the development of voice recognition applications comes many unique challenges as compared to textual recognition which is more standardized and universal. With voice recognition, you have spoken words that can vary across regional dialects. Take for example, how one region like the Southern United States can speak of certain words across that entire region. Additionally, some words may be spoken at varying speeds and emphasis depending on each individual. And finally, genders will also affect how speech is made and heard. Training an AI system to understand the uniqueness of the human language is no small task. Many of the voice recognition systems used today utilized a number of algorithms and models to be successful in their applications. Among the most popular seems to be the Hidden Markov Models. The stochastic nature of HMMs makes them incredibly useful for the speech recognition process. Taking an input and then statistically providing an output is the basis of Markov Models by providing an n-dimensional valued vector. Training of HMMs is relatively simple and automatic. For each output given by the HMM, a covariance Gaussian statistical distribution is given for the likelihood of what the system believes is the correct output variable.

On the other hand, Neural Networks are vastly different from their Hidden Markov Model counterparts. While HMMs make statistical assumptions, Neural Networks do not. Neural Networks consist of Input Layers, Hidden Layers and Output Layers. Neural Networks model themselves in a similar fashion to the human brain using nodes that are similar to neurons to send signals and make output decisions based off inputs. The number of layers involved and increasing of training data and learning rates makes neural networks smarter but not always efficient. This is due to the amount of resources needed to compute this data as well as the computing time to calculate such data.

Additional algorithm methods used for speech recognition include Dynamic Time Warping (DTW) and end-to-end speech recognition. Dynamic time warping can most often be found being used in conjunction with the Hidden Markov Model as the DTW finds a non-linearly “warped” match between two speaking speeds. The DTW establishes a middle ground speed to make its best determination on what the user is actually saying based off the input speed at which they said it. End-to-end speech recognition has a unique characteristic in that it learns all of the individual components of speech to establish the overall speech recognition training set. Therefore, no further breakdown of the training data is needed, reducing training time and data, to further the training of the AI.

III. Literature Review

The first journal report presents the different types of techniques used for text summarization [1]. The key idea of text summarization is to condense the text while keeping the meaning. The challenge in text summarization is the way humans’ summaries vs how machine can. The only way for a person to summarise text is to read and understand it, while the computer lacks that understanding of what the words mean. This paper talks about two main ways for automatic text summarization, and that's by extraction and abstraction. With extraction, it just finds the key parts of the text and returns them as they are while with abstraction, it summaries using its own words. The way humans’ summaries text is by abstraction. We don't just copy and paste, we usually use our own words to create the text summary, while this is great for humans, this does not work so well for machines, and so text extractions is mainly used. The first technique in text summary via extraction is to deal with word and phrase frequency. The idea in here is to add weight to certain words, like common words would have very little weight compared to the others. The goal here to have a ranking of every sentence to be able to identify which sentence has the highest weight. The summary would just include the top few sentences with the highest weight to them.

Once the all the weights are set, one can either go the greedy approach or convert this into an optimization problem to select the sentence for the summary.

Next they talked about topic representation approaches. As it is, machines don't understand the true meaning of the words, so how would they know what the topic of the paper is? That’s where topic representations comes into play. To find the topic

of the paper they used something called the topic signature using the log likelihood ratio test. Another approach to find the topic of the paper is similar to what was discussed before. By using the weight on the worlds, a word with higher weight would indicate that it’s a key word in the paper.

Next was the term frequency-inverse document frequency. With TFIDF, it does not require a stop word list which the other technique needed. A benefit of this is that creating this list is time consuming sometimes. With TFID the way weight is given is that it checks the importance of the word and is able to detect common words by giving a lower weight to words that appear often. With these models, few issues come up; one being that each sentence is independent from each other, which is where the Bayesian topic models tries to fix. With this model, it’s able to find similarities and difference between other documents.

In the end, all these models are lacking one key feature to create more human like summaries and that’s not taking into account text semantic. The use of knowledge base like wiki might help machine understand the words better and in turn create a better abstraction algorithms.

[3], in this paper, they talk about the use of multi sentence compression to create summaries of text. The main idea in MSC is to group similar sentence together, and from that grouping, create a new sentence that is shorter than the collection. The one downside to MSC is the degrade in the grammar when it came to generating the summary. The more detailed the summary was, the worse in grammar it was. They look to solve this issue by enhancing the way MSC works more specifically the way the word graph is made, and enhance by pre-processing. Word graphs consist of a list of stop words and the ability to tag if a word is a noun, verb, etc. They do this by first using the Multiword Expressions, which combines similar meaning phrases into one. An example that was used in the article was “to kick the bucket” which in the context means to die, and not the action of kicking a bucket. This phrase would be combined with other phrases that mean to die. Next step is similar to the first step, but with words. The goal is to find a single word to replace all the phrases, for example to use “to die” instead of “to kick the bucket” and also combine words that are synonyms when creating the word graph. The last step that was taken was to use a POS-based Language Model for grammaticality ranking of each compress sentence generated. In the end, this model provided an overall improvement in the summary of the text compared to the baseline models.

# Methodology

Improving the way something like Google home works will primary consist of its ability to summarize online text. In [1], the team at Google has provided an open source code that is able to summarize text using TensorFlow. They go on to talk about how there are two ways of going about summarizing the text. The first is to extract keywords from the text and just combine them as a summary. The next approach seems better in that it summaries in a more human like way. They use “abstractive summarization” which simply rephrases the text to what seems to be a more basic structure. The core of this technique is the “Sequence to Sequence Learning with Neural Networks” which reference [2]. The main model used to achieve this is the Long Short Term Memory Recurrent Neural Networks. The advantage of the LSTM RNN is the ability to take into consideration what it learned in the past and apply it with new data to learn about the future. This is key for something like text summarizing, to be able to predict what the next word should be and to drop words that are seen as useless later in time. The major downside to this is that it takes a long time to train the model. The team at google used a few computers and took around a week to train the model. The results some the model provided great human like summary, but the time it took is not practical. With that, other methods were explored that provided much faster results but at a cost of how well the summary is formed.

# implementation

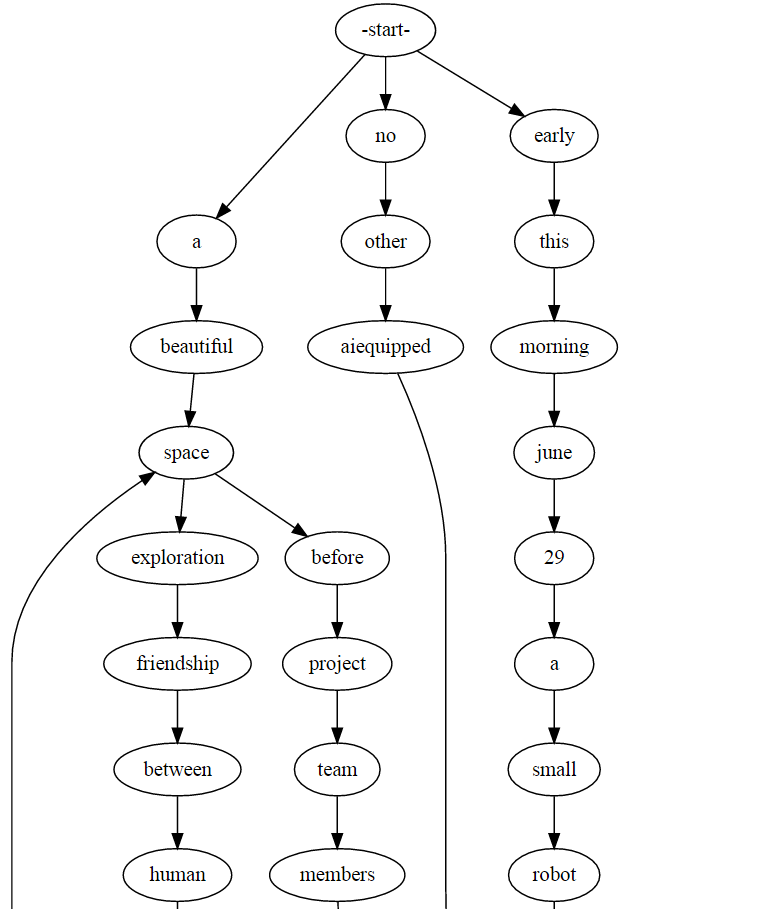
The idea we had was not to focus on the pre-processing, but to implement something after the summary has been generated. With post processing, a grammar checker was implemented using the available sources out there. This removes the focus on creating complex word graphs, but instead just passes everything into the grammar checker to fix it. To get started, we used Takahe class that is available online [8]. Takahe provided the bases in text summarizes with the ability to compress text. Takahe does two main things; first it creates a word graph with all the words that are in the text. This word graph is a graph of all the words that shows how they are linked to each other. Each sentence is added to the graph one by one until the whole paper is represented by this graph. In the end, you have something like a crossword puzzle where sentences are embedded in this graph. The summery generation works by finding the shortest path in the graph which each word has some type of weight to it. More frequent the word is, the higher weight they have. This weight is important since it helps the computer decide what words it should use for the generated summary. From this, few summaries are generated, each with their own ranking. The higher the ranking the better the summary is in terms of information importance. In Takahe, there is one important parameter that can impact the way a summary is generated. The “nb\_words” parameter creates a minimum number of words that a summary must have. The trick is trying to find the perfect number. If the number is too low, the summary would be less informative, too high; the summary would be more grammatically incorrect.

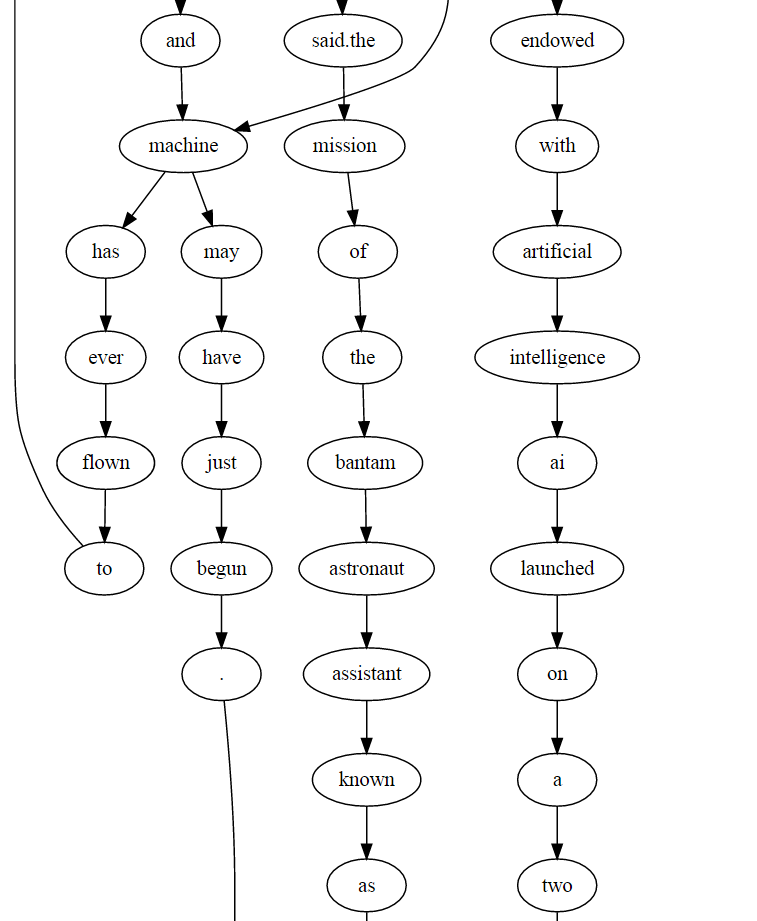
Next, the nltk library was added [6]. This library is a popular platform for Natural Language. It provided a wide range of tools from tokenize words for creating summaries of text. The text summarization this library provides works differently then Takahe. The summary generated is less human like since all it does is combine high ranking sentence together as the summary. This is clearly is not how a human would do it since summaries by humans are written in their own words. Tech differs by its ability to combine parts of sentences and formulate a completely new sentence. For this application, only the “word\_tokenize” function is used from nltk. The reason for word tokenization is it adds a tag to each word that represent if it’s a noun, a verb, etc. This automation of word tokenization is not built into Takahe which would require a manual process without nltk library. This format is key for Takahe to work properly and to build a word graph to generate a summary.

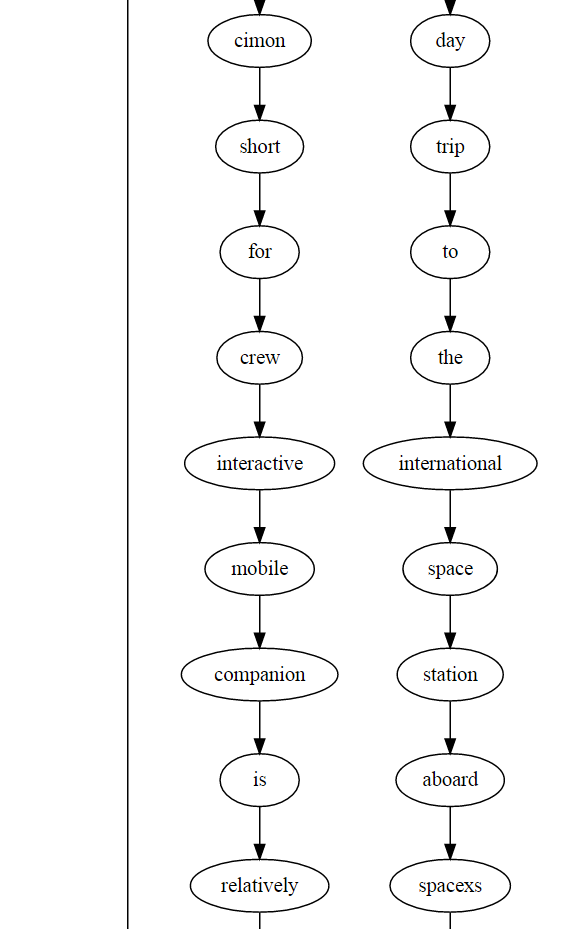
The last library that was used was ginger [7]. This is where the post processing comes into play. Ginger is an API that provides access to “getginger” grammar check website. Using the provided code, one is able to send a string of text to the api and in return it provides spelling and grammar fixes if needed.

# Results

The article that was chosen to test this text summation was about the first AI robot to operate in space [10]. The first thing that had to be done with the text was to remove all the special characters. Characters like brackets cause issues with the application, so a manual remove of them had to take place. Once that was done, all that needed to be was to send the text into the model. The results from this article were rather disappointing. When sending in the whole article, it would generate a summary that had no informative aspect to it. Only when sending a single paragraph of the article did it provide better result. Even with this single paragraph, the summary was still rather disappointing compared to the summary of the sample provided. Table 1 has results of four different texts. Each text was executed three times where each test had a different value for the minimal word count. In the end, the goal was to improve on the grammar, and not the informative aspect of the summary, and that’s where improvement was seen. Grammar was improved on most of the summaries that were sent into the model. The Grammar post processing was not perfect since it would sometimes think a word is misspelled when in fact it’s a name. The first three rows in table 1 were the sample text proved with Takahe. The summary produced from the sample text provided a lot of details with only minor grammar fix required. The fourth to sixth row summaries are generated using the article, [11]. For this, the whole article was sent in, while for row seventh to ninth, the first paragraph was sent in. The last set of summaries was chosen to see if different topic would impact the summary, but in the end that didn’t seem the case. With this, text summarization still has a long way to go to producer human like summaries, which might be a good thing. It would be interesting to see how formal schooling would adapt to this powerful tool once it matures. Having students read a book and provide a summary would become more difficult to prevent students from cheating and just having AI summarize it. As with all emerging technologies, it provides a positive impact, but also a negative impact with people exploiting it.







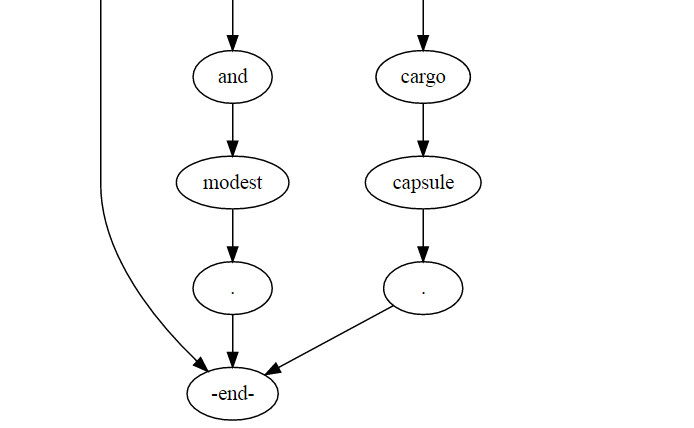


FIG 1. Word graph generated from sentences.

|  |  |  |
| --- | --- | --- |
| Min Word count | Original text | Grammar post processing |
| 1 | the wife of a former u.s. president bill clinton hillary clinton visited chinese officials . | The wife of a former u.s. president bill Clinton, Hillary Clinton visited Chinese officials. |
| 10 | the wife of a former u.s. president bill clinton hillary clinton visited chinese officials . | The wife of a former u.s. president bill Clinton, Hillary Clinton visited Chinese officials. |
| 20 | the wife of a former u.s. president bill clinton hillary clinton paid a visit to the people republic of china on monday . | The wife of a former u.s. president bill Clinton, Hillary Clinton paid a visit to the peoples republic of china on Monday. |
| 1 | it will have schulien added . | It will have scaling added. |
| 10 | we will fly towards alexander gerst might have schulien added . | We will fly towards Alexander Gerst might have stolen added. |
| 20 | we will be able to fly towards alexander gerst who arrived at the iss cimon project leader at dlr said . | We will be able to fly towards Alexander Gerst, who arrived at the is common project leader at the side. |
| 1 | a beautiful space before project team members said . | A beautiful space before project team members said. |
| 10 | a beautiful space exploration friendship between human and machine may have just begun . | A beautiful space exploration, friendship between human and machine may have just begun. |
| 20 | the mission of the bantam astronaut assistant known as cimon short for crew interactive mobile companion is relatively short and modest . | The mission of the bantam astronaut assistant known as simian short for crew, interactive mobile companion is relatively short and modest. |
| 1 | mr trudeaus decision to retaliate won a rare endorsement from all three of canadas major political parties . | Mr Trudeau's decision to retaliate won a rare endorsement from all three of Canada s major political parties. |
| 10 | mr trudeaus decision to retaliate won a rare endorsement from all three of canadas major political parties . | Mr Trudeau's decision to retaliate won a rare endorsement from all three of Canada s major political parties. |
| 20 | on social media they are calling for boycotts of american products and encouraging one another to look elsewhere for vacation destinations . | On social media they are calling for boycotts of American products and encouraging one another to look elsewhere for vacation destinations. |

Table1.- Result from sentence summary with grammar check.

References

[1] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. Tripper, J. Gutierrez and K. Kochut, "Text summarization techniques: A brief survey," *arXiv,* 2017.

[2] Liu , Peter, and Xin Pan. “Text Summarization with TensorFlow.” *Google AI Blog*, 24 Aug. 2016, ai.googleblog.com/2016/08/text-summarization-with-tensorflow.html.

[3] E. ShafieiBavani, M. Ebrahimi, R. Wong and F. Chen, "An efficient approach for Multi-Sentence compression," in *JMLR: Workshop and Conference Proceedings*, 2016.

[4] Sutskever, Ilya, et al. “Sequence to Sequence Learning with Neural Networks.” *Cornell University Library*, 14 Dec. 2014, arxiv.org/abs/1409.3215.

|  |
| --- |
| [5]C. Boyd, “The Past, Present and Future of speech recognition technology,” *Medium,* pp.https://medium.com/swlh/the-past-present-and-future-of-speech-recognition-technology-cf13c179aaf, 10 Janurary 2018.  [6]“Natural Language Toolkit.” *Nltk*, [www.nltk.org/](http://www.nltk.org/).  [7]Zoncoen. “Zoncoen/Python-Ginger.” *GitHub*, github.com/zoncoen/python-ginger.  [8]“boudinfl/takahe.” GitHub,github.com/boudinfl/takahe  [9]Bird,, Steven, et al. “5. Categorizing and Tagging Words.” *Preface*, www.nltk.org/book/ch05.html.  [10]Wall, Mike. “Meet CIMON, the 1st Robot with Artificial Intelligence to Fly in Space.” *Space.com*, Space.com, 29 June 2018, [www.space.com/41041-artificial-intelligence-cimon-space-exploration.html](http://www.space.com/41041-artificial-intelligence-cimon-space-exploration.html).  [11]Austen, Ian. “Trade War and Canadian Pride Mix in Retaliatory Tariffs Against U.S.” The New York Times, The New York Times, 30 June 2018, [www.nytimes.com/2018/06/30/business/canada-day-tariffs-trade.html](http://www.nytimes.com/2018/06/30/business/canada-day-tariffs-trade.html).  [12]Filippova, Katja. “Multi-Sentence Compression: Finding Shortest Paths in Word Graphs.” *Googleusercontent*, static.googleusercontent.com/media/research.google.com/en//pubs/archive/36338.pdf. |