Creating a Chatbot AI using Customer Support via Twitter

Jessy Francisco
California State University of Los Angeles
5151 State University Dr, Los Angeles, CA 90032
jfranc90@calstatela.edu

Deep Patel
California State University of Los Angeles
5151 State University Dr, Los Angeles, CA 90032
dpatel112@calstatela.edu

Yatri Thoria
California State University of Los Angeles
5151 State University Dr, Los Angeles, CA 90032
ythoria@calstatela.edu

Lisa Patel
California State University of Los Angeles
5151 State University Dr, Los Angeles, CA 90032
lpatel4@calstatela.edu

Rawad Moussa
California State University of Los Angeles
5151 State University Dr, Los Angeles, CA 90032
rmoussa@calstatela.edu

Abstract

This report presents the development of a chatbot AI for customer support using Twitter. The objective was to solve the problem of providing efficient and personalized customer support on a large scale. The current practice of customer support is often limited by the availability of human agents and the time it takes to respond to customer inquiries. The proposed chatbot AI aims to address these limitations by automating the support process and providing quick and accurate responses to customer queries.

The data used for training the chatbot AI includes a large collection of customer interactions on Twitter, including tweets and replies. This data provides valuable insights into the types of inquiries and issues customers commonly face. By analyzing this data, the chatbot AI can learn to understand and respond to customer queries effectively.

The approach involved using natural language processing techniques and machine learning algorithms to train a chatbot model. The model was trained on the Twitter data, leveraging pre-existing frameworks and tools for natural language understanding and dialogue generation. Several experiments were conducted to measure the success of the chatbot AI, including evaluating its accuracy, response time, and customer satisfaction.

1. Introduction/Background/Motivation

(5 points) What did you try to do? What problem did you try to solve? Articulate your objectives using absolutely no jargon.

In recent years, the boom of technology has sparked an increasingly noticeable improvement in Machine Learning (ML) and Artificial Intelligence (AI). One area in particular, Natural Language Processing (NLP), has increasingly improved the efficiency of AI chatbots in terms of understanding and response-time in concern to customer issues. As a result, there has also been a significant rise in the usage of chatbots for customer support in various industries and fields, which can be attributed to their cost-effectiveness, the assistance they provide, and the way they enhance the overall customer experience. The objective of this research project was to create our very own chatbot that aims to do exactly what we have mentioned so far, that is, respond to a user in the current context, create a personalized experience for each customer, act as around-the-clock customer service representatives, and consistency in the responses of the chatbot.

(5 points) How is it done today, and what are the limits of current practice?

Modern day chatbots are created via a combination of NLP and Machine Learning techniques. NLP techniques aid the chatbots in identifying the intent behind the customer queries, extracting key information, and determining the appropriate response. In addition, chatbots are trained using large datasets of labeled customer interactions, such as tweets. ML algorithms are employed to learn patterns and associations between customer queries and corresponding responses, which helps improve the chatbots' accuracy and effectiveness over time. There are other implementations that aid in the creation of modern chatbots, such as predefined conversational flows or decision trees, and knowledge bases, Frequently Asked Questions (FAQs), and backend systems that retrieve relevant information to the chatbots. Chatbots, however, do not come without limitations. As of today, there are many problems chatbots face that hurt their effectiveness and the customer experience, such as their lack of emotional intelligence, limited domain knowledge, difficulty in handling different languages simultaneously, and much more.

(5 points) Who cares? If you are successful, what difference will it make?

Chatbots have changed the way customer support is handled by various companies and industries. Foregone are the days where customer support is only from 8am to 5pm, or where you are put on hold for minutes on end due to a shortage of customer support assistants. With proving successful in this research project, we can show that chatbots aid and dramatically increase the performance and efficiency of customer support teams, but also reduce costs. We will do so by creating a chatbot that can reply to customer queries within a couple of seconds, and with accurate responses.

(5 points) What data did you use? Provide details about your data, specifically choose the most important aspects of your data mentioned here. You don't have to choose all of them, just the most relevant. The data used for this research project comes from a dataset found on a website named Kaggle, the central hub and community for Data Science and its surrounding fields. Specifically, this dataset is called Customer Support on Twitter, which is a collection of over 3 million tweets, composed of customer queries to some of the biggest brands on Twitter (i.e. Spotify, Walmart, Delta, etc.) and response to said customers via customer support agents. The dataset is in the form of a CSV file, where each tweet acts a row. There are various columns that help organize the data. The important columns consist of author_id, inbound, text, response_tweet_id, and in_response_to_tweet_id. The author_id and text column make up the main parts of the data, acting as the tweet and name of the anonymous user/customer. The inbound column determines whether the tweet is directed at a particular company with a Twitter. The last two columns are in correlation to any tweets responding to a main customer tweet.

2. Approach

(10 points) What did you do exactly? How did you solve the problem? Why did you think it would be successful? Is anything new in your approach? The first step to beginning this research project was plain and simple, answering the question that is, what exactly is a chatbot and how does it work.

Surely enough, various resources were collected and studied to come up with the best idea of a chatbot. Based on our understanding, chatbots process data presented to them via natural language processing through a process known as parsing. Parsing, in of itself is the process of analyzing the syntactic structure of sentences/texts based on predefined grammatical rules. In this way, we break down a sentence or text into its primal parts and determine how said parts are related to each other.

In addition to this, we found out that chatbots are generally grouped into two categories, Retrieval-based models and generative models. Retrieval-based models rely on lookup tables or a knowledge base to select an answer from a predefined set of answers. Although this method might seem naive, most chatbots in production are of this kind. Of course, there can be various degrees of sophistication with regard to selecting the best answer from the lookup tables or knowledge base. Generative models, on the other hand, generate responses on the fly instead of adopting a lookup-based approach. They are mostly probabilistic models or models based on machine learning. Considering that Generative chatbot models are more data expensive and more prone to grammatical mistakes, we went with the Retrieval-based chat model.

Once we had gotten a pretty good understanding how a chatbot works, we proceeded to the next step in our problem which was finding a model well suited for our chatbot. We ended up going with a sequence to sequence model using the Recurrent Neural Network, LSTM. The reason for choosing this architecture simply fell down to how well-suited this model architecture is for capturing the context of the customer input and then generating appropriate responses based on that. There were, however some slight modifications to the usual mode, which can be seen in the figure below:

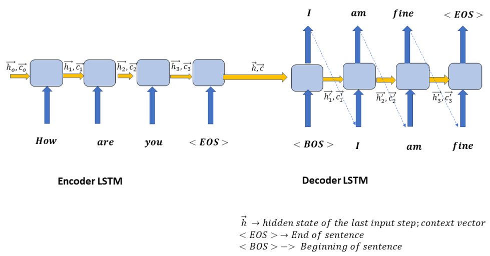

figure 2.1

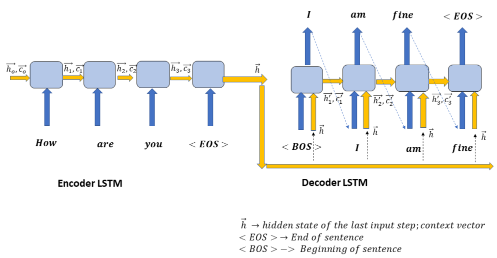

figure 2.2

Both figures depict a sequence-to-sequence model using an LSTM. However, instead of feeding the hidden state vector, h, and the cell state vector, c of the last step of the Encoder LSTM to the initial hidden an cell states of the Decoder LSTM, we feed the hidden state h at each input step of the decoder. That way, we can predict the target word at any step, the inputs are the previous target word - 1, at any step - 1, and the hidden state h. As for why we cannot go with figure 2.1, during training, for each target we know the previous word that is an input the the Decoder LSTM, However, during testing, we wouldn't have these target words, and therefore we would have to feed the previous step as an input.

From here on we went through the usual processing the data for the chatbot. We filtered customer tweets, customer service response tweets and them in correlation to their id to group all tweet conversations together. Then, the tweets were tokenized and converted into numbers so that they can be fed to our neural network and replaced anonymous screen names. And finally, we built our model for our LSTM. For our model we have two LSTMs, one for encoding the input tweet into a context vector and a second to work as a decoder for the context vector h created by the encoder. The TimeDistributed function from tensorflow.keras allows for an efficient implementation of getting a prediction at each time step of the decoder LSTM. In addition, the model is trained on categorical cross-entropy loss to predict the target words in each time step of the decoder LSTM. the categorical cross-entropy loss in any step would be over all the words of the vocabulary, and can be represented as the equation below:

$$C_t = -\sum_{i=1}^{N} y_i \log p_i$$

The label $[y_i]_{i=1}^{N}$ represents the one hot-encoded version of the target word. Only the label corresponding to the actual word would be one, while the rest would be zero. The term Pi represents the probability that the actual target word is the word indexed by i . To get the total loss, C, for each input/output tweet pair, we need to sum up the losses over all the time steps of the decoder LSTM. Since the vocabulary size might get very large, creating a one hot-encoded vector for the target label in each time step would be costly. The sparse_categorical_crossentropy loss becomes very beneficial here, since we don't need to convert the target word into a one hot-encoded vector, but instead we can just feed the index of the target word as the target label.We trained the model with the Adam optimizer due to reliable, stable convergence, along with a clip_value of 0.5, just in case the gradients become too large.

With all this finally completed, we created our training function for our model. Once trained, we used the model to generate response given an input tweet. Lastly, we put all of code together within a main function with two flows; one for training and the other for inference. (5 points) What problems did you anticipate? What problems did you encounter? Did the very first thing you tried work? The major problem we anticipated coming into this research project was a lack of knowledge in regards to chatbots, as well as knowing what libraries and modules to use for them. Thankfully there were many projects and resource throughout the web that provided stepping stones for this project, and allowed us to overcome this hurdle. In addition, however, We did not anticipate the time it would take to do so, in part to lack of time outside of this research project.

The most unanticipated problem we found during the project is the lack of power in regards to our personal computers. The first time we had completed our models and tested them, the time in regards to completing the epochs was astounding, with about 30 minutes to an hour per epoch, which in total amounted to over 3 to 4 days of letting our computers train the model. Due to this, testing and fixing our code was very difficult.

There was one minor issue that had us stumped for quiet a while, which funny enough was the most simple. Specifically, it was the inability to decode an numpy array of integers. I took about a day or two of reading documentation and online help, and the method of decoding was found.

Unfortunately, there was one major error at the end that we were unable to complete due to time constraints. Which was fully testing the model, as once again we did not have the power to fully test it in long runs. **Important: Mention any code repositories (with citations) or other sources that you used, and specifically what changes you made to them for your project.**

3. Experiments and Results

(10 points) How did you measure success? What experiments were used? What were the results, both quantitative and qualitative? Did you succeed? Did you fail? Why? Justify your reasons with arguments supported by evidence and data. To run and test the model we created, we produced several arguments via the command line. One specific example we used here case be seen below:

```
(tf2.10) C:\Users\Jessy Francisco\Desktop\nlp_project>python TwitterBotV2.py --max_vocab_size 50000 --max_seq_len 30 --embedding_dim 100
-hidden_state_dim 100 --epochs 80 --batch_size 128 --learning_rate 1e-4 --data_path "C:\Users\Jessy Francisco\Desktop\nlp_project\twcs\t
s.csv" --outpath "C:\Users\Jessy Francisco\Desktop\nlp_project"  --dropout 0.3 --mode train --num_train_records 50000 --version v1
C:\ProgramData\miniconda3\envs\tf2.10\lib\site-packages\sklearn\feature_extraction\text.py:528: UserWarning: The parameter 'token_patter
```

The entire objective of this research project was to create a chatbot that return a proper response to the user that is effective and efficient. After training our model, we were partly successful in doing so, with a few examples of customer tweets aligned with proper customer support response by the chatbot we modeled. Below are some examples of our results:

```
 able to call my mobile. Please help.']
                                              Tweet in                                          Tweet out
0  @__cname__ Oi amiga linda, vai ter mais promoç...  @__cname__ olá , UNK ! você já sabe qual conte...
1  @GWRHelp 16:45 London to Swansea very late tod...  @__cname__ hi there . sorry to hear this . can...
2  Tamagotchi Blue price is now at $ 24.43 at Ama...  @__cname__ @__cname__ i have a reliance phone ...
3  @sizehelpteam order #83977882 I ordered size c...  @__cname__ hi there , can you please dm us you...
4  @SW_Help tut tut tut.... lol 1 minute early on...  @__cname__ hi there , sorry to hear this . can...
5  Taking Lyft in St. Louis is the absolute worst...  @__cname__ we can help . please send us a dm w...
6  @__cname__ The link for the Neptune documentat...  @__cname__ hi there ! please send us a dm with...
7                      Chase Bank you owe my $12 btw.  @__cname__ hi there ! please send us a dm with...
8  Does anyone EVER get Apple's 'reading list' to...  @__cname__ hey there ! we are here to help . p...
9  @idea_cares Dear sir, my data is disabled from...  @__cname__ here to help ! send us a note via h...
Processing finished, time taken is %s 5867.636392354965
```

As shown, the model was able to accurately provide well written responses to the customer queries. However, what can also be seen is somewhat lackluster responses in terms of non-english tweets. Furthermore, time constraints and the time it took to train the model and test it, we were not able to fully test the model to its best and refine it. **Important: This section should be rigorous and thorough. Present detailed information about the decisions you made, why you made them, and any evidence/experimentation to back them up. This is especially true if you leveraged existing architectures, pre-trained models, and code (i.e. do not just show results of fine-tuning a pre-trained model without any analysis, claims/evidence, and conclusions, as that tends to not make a strong project).**

4. Other Sections

You are welcome to introduce additional sections or subsections, if required, to address the following questions in detail.

(5 points) Appropriate use of figures / tables / visualizations. Are the ideas presented with appropriate illustration? Are the results presented clearly; are the important differences illustrated?

(5 points) Overall clarity. Is the manuscript self-contained? Can a peer who has also taken Deep Learning to understand all of the points addressed above? Is sufficient detail provided?

(5 points) Finally, points will be distributed based on your understanding of how your project relates to Deep Learning. Here are some questions to think about:

What was the structure of your problem? How did the structure of your model reflect the structure of your problem?

What parts of your model had learned parameters (e.g., convolution layers) and what parts did not (e.g., post-processing classifier probabilities into decisions)?

What representations of input and output did the neural network expect? How was the data pre/post-processed? What was the loss function?

Did the model overfit? How well did the approach generalize?

What hyperparameters did the model have? How were they chosen? How did they affect performance? What optimizer was used?

What Natural Language Processing framework did you use?

What existing code or models did you start with and what did those starting points provide?

Briefly discuss potential future work that the research community could focus on to make improvements in the direction of your project's topic.

5. Work Division

Please add a section on the delegation of work among team members at the end of the report, in the form of a table and paragraph description. This and references do **NOT** count towards your page limit. An example has been provided in Table 1.

Student Name	Contributed Aspects	Details
Jessy Francisco	Data Creation and Implementation	Scraped the dataset for this project and trained the CNN of the encoder and decoder. Implemented attention mechanism to improve results. Basically everything in concerns to the project
Deep Patel	Train and test ML Model	Analyzed the dataset and trained the ML model. Implemented classifier to improve results based on tweets provided.

Table 1. Contributions of team members.

6. Miscellaneous Information

The rest of the information in this format template has been adapted from CVPR 2020 and provides guidelines on the lower-level specifications regarding the paper's format.

6.1. Language

All manuscripts must be in English.

6.2. Paper length

Papers, excluding the references section, must be no longer than six pages in length. The references section will not be included in the page count, and there is no limit on the length of the references section. For example, a paper of six pages with two pages of references would have a total length of 8 pages.

6.3. The ruler

The LaTeX style defines a printed ruler which should be present in the version submitted for review. The ruler is provided in order that reviewers may comment on particular lines in the paper without circumlocution. If you are preparing a document using a non-LaTeX document preparation system, please arrange for an equivalent ruler to appear on the final output pages. The presence or absence of the ruler should not change the appearance of any other content on the page. The camera ready copy should not contain a ruler. (LaTeX users may uncomment the `\cvprfinalcopy` command in the document preamble.) Reviewers: note that the ruler measurements do not align well with lines in the paper — this turns out to be very difficult to do well when the paper contains many figures and equations, and, when done, looks ugly. Just use fractional references (e.g. this line is 095.5), although in most cases one would expect that the approximate location will be adequate.

6.4. Mathematics

Please number all of your sections and displayed equations. It is important for readers to be able to refer to any particular equation. Just because you didn't refer to it in the text doesn't mean some future reader might not need to refer to it. It is cumbersome to have to use circumlocutions like "the equation second from the top of page 3 column 1". (Note that the ruler will not be present in the final copy, so is not an alternative to equation numbers). All authors will benefit from reading Mermin's description of how to write mathematics: http://www.pamitc.org/documents/mermin.pdf.

Finally, you may feel you need to tell the reader that more details can be found elsewhere, and refer them to a technical report. For conference submissions, the paper must stand on its own, and not *require* the reviewer to go to a techreport for further details. Thus, you may say in the body of the paper "further details may be found in [**?**]". Then submit the techreport as additional material. Again, you may not assume the reviewers will read this material.

Sometimes your paper is about a problem which you tested using a tool which is widely known to be restricted to a single institution. For example, let's say it's 1969, you have solved a key problem on the Apollo lander, and you believe that the CVPR70 audience would like to hear about your solution. The work is a development of your celebrated 1968 paper entitled "Zero-g frobnication: How being the only people in the world with access to the Apollo lander source code makes us a wow at parties", by Zeus *et al*.

You can handle this paper like any other. Don't write "We show how to improve our previous work [Anonymous, 1968]. This time we tested the algorithm on a lunar lander [name of lander removed for blind review]". That would be silly, and would immediately identify the authors. Instead write the following:

> We describe a system for zero-g frobnication. This system is new because it handles the following cases: A, B. Previous systems [Zeus et al. 1968] didn't handle case B properly. Ours handles it by including a foo term in the bar integral.
>
> ...
>
> The proposed system was integrated with the Apollo lunar lander, and went all the way to the moon, don't you know. It displayed the following behaviours which show how well we solved cases

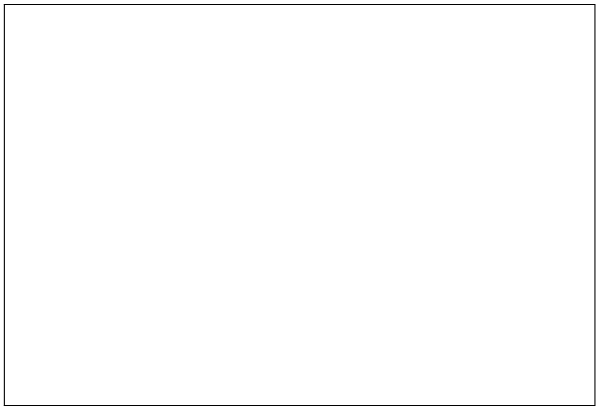

Figure 1. Example of caption. It is set in Roman so that mathematics (always set in Roman: $B \sin A = A \sin B$) may be included without an ugly clash.

A and B: ...

As you can see, the above text follows standard scientific convention, reads better than the first version, and does not explicitly name you as the authors. A reviewer might think it likely that the new paper was written by Zeus *et al.*, but cannot make any decision based on that guess. He or she would have to be sure that no other authors could have been contracted to solve problem B.

FAQ

Q: Are acknowledgements OK?
A: No. Leave them for the final copy.

Q: How do I cite my results reported in open challenges?
A: To conform with the double blind review policy, you can report results of other challenge participants together with your results in your paper. For your results, however, you should not identify yourself and should not mention your participation in the challenge. Instead present your results referring to the method proposed in your paper and draw conclusions based on the experimental comparison to other results.

6.5. Miscellaneous

Compare the following:

`$conf_a$`	$conf_a$
`$\mathit{conf}_a$`	conf_a

See The TEXbook, p165.

The space after *e.g.*, meaning "for example", should not be a sentence-ending space. So *e.g.* is correct, *e.g.* is not. The provided `\eg` macro takes care of this.

When citing a multi-author paper, you may save space by using "et alia", shortened to "*et al.*" (not "*et. al.*" as "*et*" is a complete word.) However, use it only when there are three or more authors. Thus, the following is correct: " Frobnication has been trendy lately. It was introduced by Alpher [**?**], and subsequently developed by Alpher and Fotheringham-Smythe [**?**], and Alpher *et al.* [**?**]."

This is incorrect: "... subsequently developed by Alpher *et al.* [**?**] ..." because reference [**?**] has just two authors. If you use the `\etal` macro provided, then you need not worry about double periods when used at the end of a sentence as in Alpher *et al.*

For this citation style, keep multiple citations in numerical (not chronological) order, so prefer [**?**, **?**, **?**] to [**?**, **?**, **?**].

6.6. Formatting your paper

All text must be in a two-column format. The total allowable width of the text area is $6\frac{7}{8}$ inches (17.5 cm) wide by $8\frac{7}{8}$ inches (22.54 cm) high. Columns are to be $3\frac{1}{4}$ inches (8.25 cm) wide, with a $\frac{5}{16}$ inch (0.8 cm) space between them. The main title (on the first page) should begin 1.0 inch (2.54 cm) from the top edge of the page. The second and following pages should begin 1.0 inch (2.54 cm) from the top edge. On all pages, the bottom margin should be 1-1/8 inches (2.86 cm) from the bottom edge of the page for 8.5×11-inch paper; for A4 paper, approximately 1-5/8 inches (4.13 cm) from the bottom edge of the page.

6.7. Margins and page numbering

All printed material, including text, illustrations, and charts, must be kept within a print area 6-7/8 inches (17.5 cm) wide by 8-7/8 inches (22.54 cm) high.

6.8. Type-style and fonts

Wherever Times is specified, Times Roman may also be used. If neither is available on your word processor, please use the font closest in appearance to Times to which you have access.

MAIN TITLE. Center the title 1-3/8 inches (3.49 cm) from the top edge of the first page. The title should be in Times 14-point, boldface type. Capitalize the first letter of nouns, pronouns, verbs, adjectives, and adverbs; do not capitalize articles, coordinate conjunctions, or prepositions (unless the title begins with such a word). Leave two blank lines after the title.

AUTHOR NAME(s) and AFFILIATION(s) are to be centered beneath the title and printed in Times 12-point, non-boldface type. This information is to be followed by two blank lines.

The ABSTRACT and MAIN TEXT are to be in a two-column format.

MAIN TEXT. Type main text in 10-point Times, single-spaced. Do NOT use double-spacing. All paragraphs should be indented 1 pica (approx. 1/6 inch or 0.422 cm). Make sure your text is fully justified—that is, flush left and

Figure 2. Example of a short caption, which should be centered.

flush right. Please do not place any additional blank lines between paragraphs.

Figure and table captions should be 9-point Roman type as in Figures 1 and 2. Short captions should be centred.

Callouts should be 9-point Helvetica, non-boldface type. Initially capitalize only the first word of section titles and first-, second-, and third-order headings.

FIRST-ORDER HEADINGS. (For example, **1. Introduction**) should be Times 12-point boldface, initially capitalized, flush left, with one blank line before, and one blank line after.

SECOND-ORDER HEADINGS. (For example, **1.1. Database elements**) should be Times 11-point boldface, initially capitalized, flush left, with one blank line before, and one after. If you require a third-order heading (we discourage it), use 10-point Times, boldface, initially capitalized, flush left, preceded by one blank line, followed by a period and your text on the same line.

6.9. Footnotes

Please use footnotes[1] sparingly. Indeed, try to avoid footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence). If you wish to use a footnote, place it at the bottom of the column on the page on which it is referenced. Use Times 8-point type, single-spaced.

6.10. References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example [**?**]. Where appropriate, include the name(s) of editors of referenced books.

[1] This is what a footnote looks like. It often distracts the reader from the main flow of the argument.

Method	Frobnability
Theirs	Frumpy
Yours	Frobbly
Ours	Makes one's heart Frob

Table 2. Results. Ours is better.

6.11. Illustrations, graphs, and photographs

All graphics should be centered. Please ensure that any point you wish to make is resolvable in a printed copy of the paper. Resize fonts in figures to match the font in the body text, and choose line widths which render effectively in print. Many readers (and reviewers), even of an electronic copy, will choose to print your paper in order to read it. You cannot insist that they do otherwise, and therefore must not assume that they can zoom in to see tiny details on a graphic.

When placing figures in LaTeX, it's almost always best to use `\includegraphics`, and to specify the figure width as a multiple of the line width as in the example below

```
\usepackage[dvips]{graphicx} ...
\includegraphics[width=0.8\linewidth]
                {myfile.eps}
```

6.12. Color

Please refer to the author guidelines on the CVPR 2020 web page for a discussion of the use of color in your document.

[3, 1, 2]

References

[1] Felix A. Gers, Jürgen Schmidhuber, and Fred Cummins. Learning to forget: Continual prediction with LSTM. *Neural Computation*, 12(10):2451–2471, oct 2000. 7

[2] Xiaojie Guo. supp1-3214832.pdf. 7

[3] Wiebke Wagner. Steven bird, ewan klein and edward loper: Natural language processing with python, analyzing text with the natural language toolkit. *Language Resources and Evaluation*, 44(4):421–424, may 2010. 7