CONVERSATIONAL ARTIFICIAL INTELLIGENCE IN NATURAL LANGUAGE PROCESSING APPLICATION WITH LIFELONG LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Conversational AI bot known as chatbot has many roles in human life today, such as answering uncomplicated questions on E-commerce website pages or even help as an assistant like Siri and Google assistant. However, the chatbot is currently limited due to insufficient knowledge available on a predetermined training dataset. So when a chatbot receives an unfamiliar question, it may be hard to understand the question. Therefore, we formed a chatbot using the lifelong learning method. Thus, that chatbot can conduct training on unfamiliar incoming data without train the model from scratch. The performance benchmark the system used is the average confidence score of the tests carried out. The system has an average confidence score of 0.80952.

1 INTRODUCTION

The development of artificial intelligence in this modern era has rapidly grown (Adamopoulou & Moussiades, 2020). This development is supported by technological developments in the ability of computers to perform calculations or computations that are much complicated. Thus, the technology found at this time is not limited to the "Mathematical" approach, but to a system or algorithm that can acquire knowledge without requiring human assistance or explicit programming (Saravanan & Sujatha, 2018).

The ability of a system to study data and determine its patterns provides a big potential to facilitate humans in assessing the quality of data and perform a treatment on the data to improve its quality (Wang & Tao, 2016). Therefore, the studies of machine learning are split into more specific branches. Examples of these branches such as search engines, computer vision, natural language processing, etc.

A conversational bot or what is generally called a Chatbot is an example of an application of Machine Learning in the field of Natural Language Processing (Ranoliya et al., 2017) (Liddy, 2001). This system has an embedded knowledge. So, it can identify the given sentences and then make a decision about what is appropriate to be a response to the input sentence (Setiaji & Wibowo, 2016). Currently, Chatbots can be applied to various sectors such as customer service, robotic research, or assistants who assist in household activities. Examples of chatbots used to help manage human activities are SIRI for Apple and Google Assistant for Android.

Five to ten years ago, chatbots were limited in an algorithm function. It means that chatbots can process a set of strict sentences only (Nadkarni et al., 2011). But at this time, chatbots, in general, are already using Neural Networks. Chatbots that use neural networks will initially be provided with a large amount of data/information to learn the responding procedure with grammatically relevant and accurate responses to input utterances (Vamsi et al., 2020). So, chatbot sentences used in talking to chatbots are more varied with broader capabilities and are limited to the data that has been trained. One of the network architectures used in the use of chatbots is Long Short-Term Memory used in Google Translate (Wu et al., 2016).

However, conventional chatbots with neural networks only for their architecture are limited to identify sentences based on their training data set. Thus, the chatbot will not recognize data outside of its training dataset. The direct follow-up training process for additional data will also have the potential to cause catastrophic forgetting (Mittal et al., 2021). So, transfer learning is carried out with the development of incremental learning in the development of chatbot knowledge so that the chatbot can recognize new sentences without having to ”forget” about previous knowledge.

2 PROBLEM STATEMENT

The chatbot requires an algorithm to identify the given sentence to reply to it correctly. Many algorithms can be used in chatbots, such as functions in general (Dahiya, 2017) or using neural networks. However, the drawback of all these algorithms is the chatbot can only identify sentences that have been learned. Hence, the model needs to be trained from the beginning when new sentences that foreign from the training data set occurred (Ans et al., 2004). Several observations have been made for Natural Language Processing and chatbots, namely:

1. Currently, the chatbot's ability to identify a sentence is highly dependent on algorithms and training data set. So, when new sentences that are foreign occurred, it is necessary to retrain until re-deployment. The addition of new sentences over time is crucial for chatbots. Because, the much number of new sentences that appear, the more capable chatbot can adapt to the existing sentences.
2. In a large data sets, the training process can become more complex and complicated if training is carried out from the beginning, in the sense that there is no initial load (Lomonaco et al., 2021). Moreover, the continuous addition of data will make the overall training data set become larger (He et al., 2011). The training process will become even more complex in the next training process. Thus, a ”checkpoint” is needed for the iteration's weight to facilitate or simplify the training process.
3. Continuous development of chatbots knowledge may lead to catastrophic forgetting, which means that the model's ability to recognize new words may lead to forgetting the previously trained data (Kirkpatrick et al., 2017).

From the observations above, the proposed algorithm must have the ability to learn continuously on sentences that it does not understand, without having to carry out the training process from the beginning or continue learning from the previous weight and be able to overcome the catastrophic forgetting model. So, the model can continue to develop knowledge of previous data and new data (Goodfellow et al., 2016).

3 PROPOSED SYSTEM

The chatbot on our proposed system uses a data set that we adapt to use in everyday conversations. That data set consists of $< tag >$, $< pattern >$, dan $< response >$. The category we use in this data set is the atomic category, which refers to the AIML classification or Artificial Intelligence Markup Language Classification, where the AIML classification contains sentence patterns that have been defined as a whole or precisely (Parisi et al., 2019).

$< tag >$ Asking for weather.
$< pattern >$ Do you know what weather today? $< /pattern >$
$< response >$ Sure, what city you want to know? $< /response >$

Where the tag is in the form of the sentence category or type of sentence, the pattern is a sentence that will be identified by the model or sentence inputted by the user and will later be identified or recognized by the model, and the response is a reply sentence from the chatbot to the pattern or sentence given to the chatbot. Some tags or categories may contain more than one pattern. So, the sentences entered by the user can be more varied but still within the same category or have similarities in the recognized categories.

The capability of chatbots for developing their knowledge at identifying foreign sentences as time passes by is very important (Shawar & Atwell, 2007). The chatbots knowledge development can be done by applying incremental learning to the chatbot model. When the new data is ready to be trained, the old data will be stored in a memory for the retraining process to avoid catastrophic forgetting. The flowchart of the chatbot from the start to the flow of increasing data is shown in Figure 1.

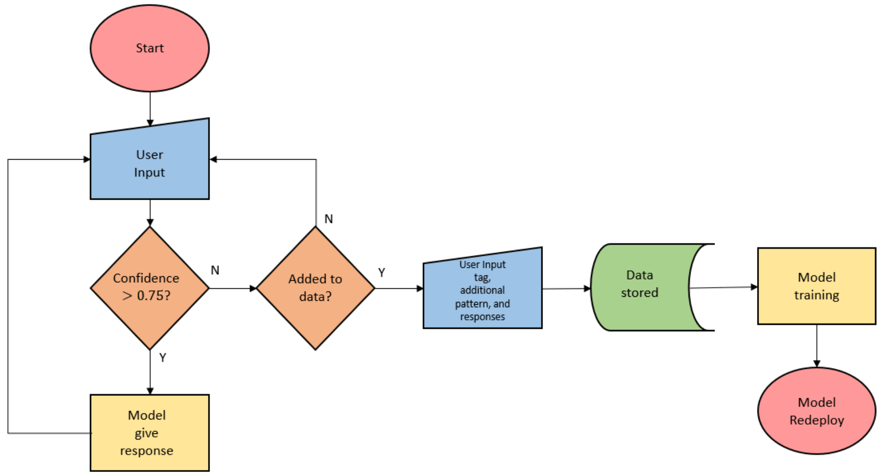
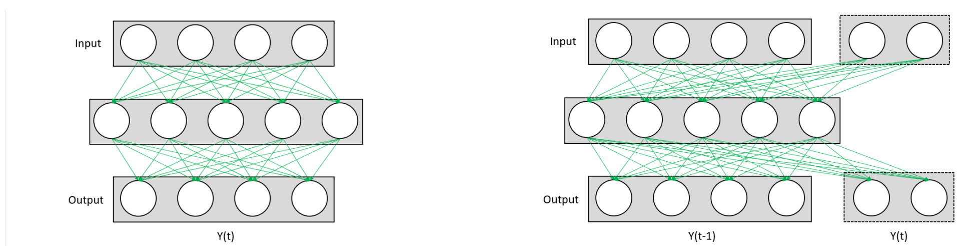


Figure 1: Flowchart chatbot.

Nevertheless, as the new categories and patterns occur, more inputs and outputs are needed as well (Albesano et al., 2006). Thus, adaptive adjustments or the development of neural network inputs and outputs are crucial, so that the system is can adapt to the addition of new categories and patterns. An example of visualization for adaptive adjustment or expansion of inputs and outputs is shown in Figure 2.



(a) Neural before data addition

(b) Neural change after data addition

Figure 2: Visualisation of input and output changes to data changes

In addition, the other key factor that plays a role in developing the chatbot's knowledge capabilities in the system is by applying transfer learning (Pan & Yang, 2009). Hence, the model does not need to do training from the start, but from the previous checkpoint or weight (Torrey & Shavlik, 2010). However, adding adaptive inputs and outputs causes transfer learning to experience a few obstacles, namely the different weights on the previous input and output. The solution is to add dummy weights to the new inputs and outputs, without changing the existed weight. So, when transfer learning happens, the training process on the model does not require too many adjustments compared to the conventional training process. The pseudocode of the algorithm is shown in Algorithm 1.

Algorithm 1 Dummy weight

```
if previous weight is exist then
    model load previous weight
    additional weight = input size-previous input size
    add additional weight equal to additional weight
    put for additional weight = 0
end if
```

The addition of a transfer learning algorithm to the model can overcome the catastrophic forgetting that occurs in the model during the application of incremental learning to achieve continuous learning in the model for chatbots (Polikar et al., 2001).

4 Result

Two types of training are applied for the training process, namely basic training and incremental learning. In the basic training method, the model is trained directly without any initial weight. For the incremental training method, the model is trained on 75% from the data set, and later the training process will be carried out again on the new data of the rest 25%. So, The number of training data set for both methods would be equal. The data set used in this training process itself uses a self-made data set and is customized for everyday conversation. The algorithm used to carry out this training process is Cross Entropy based on a neural network. Loss in training is shown in Figure 3.

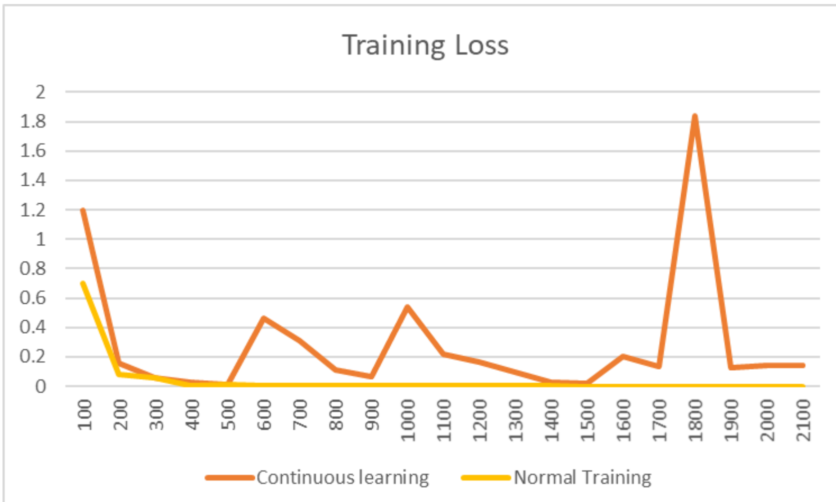


Figure 3: Training Loss

For every existing incline, it occurs because there is additional data to the model during training so that the model is still trying to adjust to both old and new data. There is a significant incline in the 1800th epoch, that's because the added data has more patterns, so the model requires more adjustments when compared to adding data as before.

To validate the model itself, the model will try to identify the sentence given to the user and later return the confidence value to the existing data. So, this confidence value will be used as a validation parameter for the model. The results of the average confidence of the model are shown in Table 1.

Table 1: Average confidence

METHOD	**AVERAGE CONFIDENCE**
Basic Training	0.81811
Continuous Training	0.80952

The average confidence from basic training is higher than the continuous training. It is because the model is better to adjust the weight on the data if the training process is executed in a single run without any significant additional data, even though the average confidence from continuous training is almost the same (Gepperth & Hammer, 2016). There are two false negatives occurrence for the basic method and two false positives for the continuous ones. For the confidence model assessment itself, if the model predicts correctly and accordingly, the calculation of the average confidence would be performed. Contrary when the model prediction is incorrect, the confidence score will be considered 0.

5 Conclusion

In this paper, we propose an algorithm that can be used to improve the chatbot data development capability by combining incremental learning capabilities by learning the data one by one and transfer learning as the ability to learn new data based on previous data to achieve continuous learning capabilities. Based on our observations and evaluations, this algorithm can run quite similar to the basic training process on a neural network with an average confidence score that is quite close. This algorithm can be used to chatbots model that will continuously learn or apply continuous learning to the algorithm. Several improvements can be made, namely using a better classification algorithm so that later you can get better confidence score results, and it is also possible to make training losses lesser than the existing ones.

References

Eleni Adamopoulou and Lefteris Moussiades. An overview of chatbot technology. In Ilias Maglogiannis, Lazaros Iliadis, and Elias Pimenidis (eds.), *Artificial Intelligence Applications and Innovations*, pp. 373–383, Cham, 2020. Springer International Publishing. ISBN 978-3-030-49186-4.

Dario Albesano, Roberto Gemello, Pietro Laface, Franco Mana, and Stefano Scanzio. Adaptation of artificial neural networks avoiding catastrophic forgetting. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, pp. 1554–1561. IEEE, 2006.

Bernard Ans, Stéphane Rousset, Robert M French, and Serban Musca. Self-refreshing memory in artificial neural networks: Learning temporal sequences without catastrophic forgetting. *Connection Science*, 16(2):71–99, 2004.

Menal Dahiya. A tool of conversation: Chatbot. *International Journal of Computer Sciences and Engineering*, 5(5):158–161, 2017.

Alexander Gepperth and Barbara Hammer. Incremental learning algorithms and applications. In *European symposium on artificial neural networks (ESANN)*, 2016.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Haibo He, Sheng Chen, Kang Li, and Xin Xu. Incremental learning from stream data. *IEEE Transactions on Neural Networks*, 22(12):1901–1914, 2011.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Elizabeth D Liddy. Natural language processing. 2001.

Vincenzo Lomonaco, Lorenzo Pellegrini, Andrea Cossu, Antonio Carta, Gabriele Graffieti, Tyler L. Hayes, Matthias De Lange, Marc Masana, Jary Pomponi, Gido van de Ven, Martin Mundt, Qi She, Keiland Cooper, Jeremy Forest, Eden Belouadah, Simone Calderara, German I. Parisi, Fabio Cuzzolin, Andreas Tolias, Simone Scardapane, Luca Antiga, Subutai Amhad, Adrian Popescu, Christopher Kanan, Joost van de Weijer, Tinne Tuytelaars, Davide Bacciu, and Davide Maltoni. Avalanche: an end-to-end library for continual learning. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2nd Continual Learning in Computer Vision Workshop, 2021.

Sudhanshu Mittal, Silvio Galesso, and Thomas Brox. Essentials for class incremental learning. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 3508–3517, 2021.

Prakash M Nadkarni, Lucila Ohno-Machado, and Wendy W Chapman. Natural language processing: an introduction. *Journal of the American Medical Informatics Association*, 18(5):544–551, 2011.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

Robi Polikar, Lalita Upda, Satish S Upda, and Vasant Honavar. Learn++: An incremental learning algorithm for supervised neural networks. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 31(4):497–508, 2001.

Bhavika R. Ranoliya, Nidhi Raghuwanshi, and Sanjay Singh. Chatbot for university related faqs. In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 1525–1530, 2017. doi: 10.1109/ICACCI.2017.8126057.

R. Saravanan and Pothula Sujatha. A state of art techniques on machine learning algorithms: A perspective of supervised learning approaches in data classification. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 945–949, 2018.

Bayu Setiaji and Ferry Wahyu Wibowo. Chatbot using a knowledge in database: Human-to-machine conversation modeling. *2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS)*, pp. 72–77, 2016.

Bayan Abu Shawar and Eric Atwell. Different measurement metrics to evaluate a chatbot system. In *Proceedings of the workshop on bridging the gap: Academic and industrial research in dialog technologies*, pp. 89–96, 2007.

Lisa Torrey and Jude Shavlik. Transfer learning. In *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*, pp. 242–264. IGI global, 2010.

G Krishna Vamsi, Akhtar Rasool, and Gaurav Hajela. Chatbot: A deep neural network based human to machine conversation model. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–7, 2020.

Jue Wang and Qing Tao. Machine learning: The state of the art. *IEEE Intelligent Systems*, 23(6): 49–55, 2016.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, ukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016.