

I. D. Nyzhnyk

VOICE RECOGNITION USING REINFORCEMENT MACHINE LEARNING MODEL

Abstract. This paper presents the method for improvement of quality of voice recognition machine learning model by using users habits and feedback for each unique case (chat). Chat bot implemented as an interface for machine learning models.

Keywords: reinforcement learning, machine learning, voice recognition, messenger, chat bot.

Introduction. Nowadays we have a bunch of solutions for recognition of voice messages in messengers. Most of them use third-party cloud services for voice recognition. This results in a non-flexible solution that tries to cover more cases at the cost of lower quality.

The goal of this paper is to prove that a custom machine learning model trained on specific cases has better performance and quality metrics.

Research results. A chatbot was created for “Telegram” messenger as an interface for machine learning models. There are two machine learning models: base model, trained with audio books and voice samples, and an extension for the base model for each unique chat that uses reinforcement learning to improve results.

The base model is basically a convolutional neural network [1]. Convolutional neural networks have had great success not only in voice recognition but also in image and signal processing [2] and are the most common architecture for such tasks. The extended model is the base model with reinforcement learning applied.

In response to each voice message in chat bot provides two results from each of models with ability to give feedback and correct the message. As we have a self extended model for each chat, feedback will be used to re-train only the model for specific chat and won't have impact on models for other chats.

In result we got a better level of accuracy compared with existing analogs. For comparison we took Voicy Bot which uses wit.ai as the most common solution for speech recognition in chats.

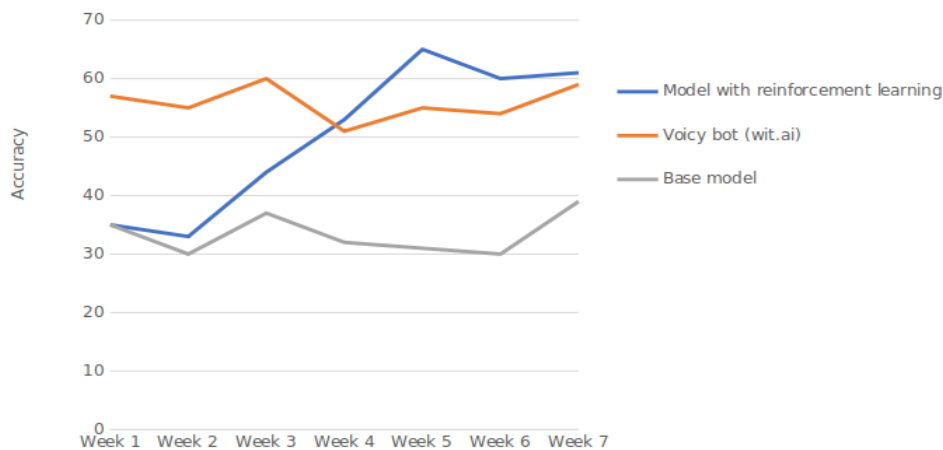


Fig. 1. Accuracy of used neural networks in time

As you can see on the chart above, because of users feedback and reinforcement learning accuracy is growing up in time, while accuracy of other models remains the same.

Conclusion. Novelty of proposed method is the use a unique machine learning model for each chat of users. It results in better accuracy in public chats. The main weakness of the model is lower accuracy outside of the chat it was trained on.

References

1. O. Abdel-Hamid, A. Mohamed, H. Jiang, L. Deng, G. Penn and D. Yu, "Convolutional Neural Networks for Speech Recognition," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 10, pp. 1533-1545, Oct. 2014, doi: 10.1109/TASLP.2014.2339736.
2. M. Kozlenko, I. Lazarovych, V. Tkachuk and V. Vialkova, "Software Demodulation of Weak Radio Signals using Convolutional Neural Network," 2020 IEEE 7th International Conference on Energy Smart Systems (ESS), Kyiv, Ukraine, 2020, pp. 339-342, doi: 10.1109/ESS50319.2020.9160035.