Self-optimization of mobile robot actions using voice commands

A I Izyumov, E V Vasilyeva, E A Alentsov and O S Ostapovich

Don State Technical University, Rostov-on-Don, Russia

Annotation. In most cases, multifunctional robots used both in industry and in the home environment to facilitate human labor have a common problem, which is a lack of autonomy and self-learning. This does not mean that it is impossible to implement self-learning within the operation of these machines, but most often such technical units function according to a specific algorithm, without deviating from its principles, even if this may lead to better results. First of all, a clear adherence to the program, in accordance with the acceptable quality of the final result, is put at the forefront. However, when working with technical units that do not depend on clear industrial or household algorithms, it is allowed to use an individual approach to the implementation of management processes. This work is devoted to the construction of the concept and principles of implementing voice control of a mobile robot for performing movement tasks.

1. Introduction
This paper presents the concept of a robot that uses neural networks to recognize voice commands. It is possible to integrate into the previously presented development of a mobile robot runner for novice athletes, competing with a group or individual athlete on running tracks with a set time when moving around the stadium [1].

In an analytical review of existing market solutions, it was found that the orientation systems of many products are most often represented by logical systems based on a variety of sensors. The proposed system is implemented as a neural network and requires only the presence of cameras and an inertial sensor of the robot's position for further orientation in space and avoiding unwanted collisions. In General, such a system is more flexible and cheaper due to the smaller number of additional sensors, and the only factor limiting the quality of the robot's orientation is the quality of the training procedure performed, taking into account the amount of data used.

As you know, a neural network is an artificial, multi-layered highly parallel (i.e., with a large number of independently parallel elements) logical structures made up of formal neurons [2].

Main applications of neural networks in mobile robotics:
- automation of the classification process;
- automation of forecasting and recognition;
- automating the decision-making process;
- managing, encoding, and decoding information.

In further work, the authors are primarily interested in the interaction of neural networks and sensor signals, the problems of signal perception and recognition in intelligent systems. Since the problem of signal recognition has been and remains relevant, more and more advanced methods of solving it are being developed [3,8,9,10].
2. Concept
If the economic aspect is not taken into account, the main problem hindering the widespread use of robotic systems is the problem of the limited scope of robot applications. Creating universal robotic systems that can be used for a wide range of tasks is still a partially unsolved problem. [12]

The main disadvantages of existing speech recognition systems are the need for a clear audio signal and a system that can separate extraneous noise from the informative part of the signal without significant losses. This problem is proposed to be solved using the technology of the silent speech interfaces (SSI). [11]

The concept is based on a robotic mobile platform with automatic workflow optimization based on reinforcement learning. [12] the robot's trajectory is formed by the operator's voice commands. After the initial implementation of the movement process, it is possible to change the trajectory based on additional adjustments.

Additionally, two cameras on the body provide the ability to avoid unwanted collisions with environmental objects. If a separate device is developed that is not integrated into the previously presented concept [4], there are mounting slots on the roof of the case, which are used, for example, for attaching an Autonomous manipulator, or a transport basket for transporting small loads. You can also add a Doppler radar if you need to work among other robots or recognize voice commands from behind a wall [6].

The appearance of the mobile transport platform is shown in Figure 1.

![Figure 1. Appearance of the mobile transport platform](image)

The location of external modules of the mobile transport platform is shown in Figure 2.

![Figure 2. Location of external modules of the mobile transport platform](image)
The software trainable part of the robot consists of 2 trainable recurrent neural networks. The first part of the neural network includes four layers, two of which are hidden. [7] The structure of a four-layer neural network is shown in figure 3. The input of the neural network receives an audio signal and falls on the first input layer of the first of the two neural networks, where the recorded audio track is divided into segments of 25 milliseconds and 10 milliseconds in increments. [19] This fragmentation is necessary so that in each segment the next hidden neural layers can detect the word among the stream of noise. After splitting over a short 25 millisecond track, it is transformed into a matrix of values, which is understandable for the neural network. [15] The resulting matrix is sent sequentially to two hidden layers, where its values are added to such pre-selected coefficients during training, so that the output layer has a specific integer symbolizing the label of a specific word. The connection between placemarks and audio tracks has already become known to the neural network after the machine learning procedure. The resulting integer values are written to a variable and based on its value, the robot moves straight, turns, or stops.

![Figure 3. The structure of four-layer neural network](image)

Test suites for neural network training are recorded under different background noise conditions. The number of input neurons is equal to the number of values of the matrix on which the audio track with the voice command is laid out. [13]

The second part of the network also consists of four layers and its main task is to adjust actions based on voice commands after the end of the action cycle. A matrix consisting of 2 rows is fed to the input layer of the second neural network. [7] This matrix represents the order of actions recorded by the robot. The first line contains the values of a variable that describes actions (stop, forward, left, right). The second line describes the instants of time in seconds in which the action takes place. An example of such a matrix is shown in Table 1. In this case, "1" indicates movement, "2" a 90-degree turn and subsequent stop, "3" a -90-degree turn and subsequent stop, and "0" a stop and end of the path. The values of the described matrix are rounded and fed to the hidden layers. The task of hidden layers is also to add a certain coefficient to the matrix values based on the previous recorded passes of one path, but this time the task of the output layer is to collect a new generated matrix of values. A new path is generated each time the previous path is completed. This adaptability can be useful in a constantly changing environment. [20]
Table 1. The order of actions recorded by the robot

|   |   |   |   |
|---|---|---|---|
| 1 | 2 | 1 | 3 | 0 |
| 3.51 | 8.03 | 42.97 | 46.15 | 48.29 |

The principle of operation is based on repeating the robot's movements. Each movement is part of a cycle in which the robot implements the necessary actions. In this way, you can quickly and efficiently retrain the robot for a new task or change an existing one without additional software. [16]

The mobility of the platform implies its active use, therefore-easy accessibility and the ability to quickly replace components. If the concept of a mobile robot designed to perform running exercises is taken as the basis [1], then components are used:

Table 2. Material costs for development

| Name of materials, Ed. ed. | Units | Quantity | Price per unit, RUB. | The total amount of |
|---------------------------|-------|----------|----------------------|---------------------|
| ARDUINO NANO Platform     | Pcs.  | 1        | 260                  | 260                 |
| Sony color analog CVBS bodyless video camera 480 TVL 0.3 Lux | Pcs.  | 2        | 450                  | 900                 |
| L293d engine driver       | Pcs.  | 1        | 220                  | 220                 |
| Planetary gear motor DC PTC7152 | Pcs.  | 2        | 1500                 | 3000                |
| Red ENERGY DS 1207.1 battery | Pcs.  | 1        | 2530                 | 2530                |
| Resistor MO-200           | Pcs.  | 4        | 6                    | 24                  |
| Housing (aluminum 3x1200x3000 mm 1105anr) | Pcs.  | 2        | 210                  | 420                 |
| Body processing           | Pcs.  | 1        | 400                  | 400                 |
| Wheel 65 mm               | Pcs.  | 4        | 81                   | 324                 |
| Total:                    |       |          |                      | 8078                |

Dimensions are not specified, as they may change directly during the assembly process.

3. Conclusion

The main advantages of the proposed concept are self-learning and ease of use without the need for additional software to manage the platform by the operator when using voice commands [14, 17].

The basic nature of the platform provides a wide range of options for replacing individual elements in the event of a malfunction, as well as reducing the risk of disabling the device. It was mentioned above that the concept can be integrated into the previously presented development of a mobile robot runner for novice athletes. [1] control of the robot is planned directly, using voice.

The project can be used both for the development of household devices (in particular, robotic home cleaners of the new generation), and in sports, and for the construction of transport vehicles.

References

[1] A. Izumov, E. Vasilyeva, S. Lobanov, A. Sergienko, E. Alentsov. Mobile Robot for Running Exercises
[2] https://bigenc.ru/technology_and_technique/text/4114009 (Electronic resource, last checked on 22.08.2020)

[3] https://cyberleninka.ru/article/n/audiovizualnyy-rechevoy-interfeys-dlya-sistem-upravleniya-i-opovescheniya

[4] https://cyberleninka.ru/article/n/problemy-formirovaniya-sistem-obespechivayuschih-avtomaticheskoе-raspoznavanie-obektov-po-dalnostnym-portretam

[5] https://cyberleninka.ru/article/n/problemy-formirovaniya-sistem-obespechivayuschih-avtomaticheskoе-raspoznavanie-obektov-po-dalnostnym-portretam

[6] https://cyberleninka.ru/article/n/postroenie-adaptivnyh-algoritmov-servoupravleniya-manipulyatorami-na-osnove-obratных-zadach-dinamiki-i-neyronnych-setey

[7] https://cyberleninka.ru/article/n/ob-obodom-algoritme-posledovatelnoy-initializatsii-vesov-glubokih-neyronnych-setey-i-obuchenii-ansamblya-neyronnych-setey

[8] https://patent.ru/patent/RU2291499C2

[9] https://patent.ru/patent/RU2643452C2

[10] https://patents.google.com/patent/US10679628B2/en

[11] https://ru.m.wikipedia.org/wiki/Интерфейс_безмолвного_доступа

[12] https://ru.m.wikipedia.org/wiki/Обучение_с_подкреплением

[13] https://ru.m.wikipedia.org/wiki/Распознавание_речи

[14] Kim Joo-Chang 2018 Neural-network based adaptive context prediction model for ambient intelligence

[15] Lery Sakti Ramba 2020 Design Of A Voice Controlled Home Automation System Using Deep Learning Convolutional Neural Network (DL-CNN) pp 57 – 73

[16] Marcus Hutter 2010 Towards a Universal Theory of Artificial Intelligence Based on Algorithmic Probability and Sequential Decisions Manno-Lugano pp 3-5

[17] Reddy, Yanala Abhimanyu 2020 A Neural Network Based Algorithm for Identifying Speaker Voice The International journal of analytical and experimental modal analysis pp 1573-1579

[18] Rohan Khanna 2019 Through-Wall Remote Human Voice Recognition Using Doppler Radar With Transfer Learning pp 491

[19] Tandel Nishtha H. 2020 Voice Recognition and Voice Comparison using Machine Learning Techniques: A Survey pp 173

[20] Yip Michael C. 2019 Neural Path Planning: Fixed Time, Near-Optimal Path Generation via Oracle Imitation