IRSTI 28.23.15 https://doi.org/10.26577/JPCSIT.2023.v1.i1.010

O.D. Nuriddinov*, Zh.Zh. Omirbekova

Al-Farabi Kazakh National University, Almaty, Kazakhstan
*e-mail: nuriddin.atabek2000@gmail.com

HAND GESTURE RECOGNITION USING MACHINE LEARNING ALGORITHMS

Abstract. Communication between people is an integral part of life. In the process of communication, people convey their emotions, thoughts, desires to each other. People with disabilities, such as deaf and dumb people, experience various difficulties in the process of communication [1]. Today, 5% of the world's people (more than 430 million), and in Kazakhstan more than 18 thousand people suffer from deafness. By 2050, this it is assumed will reach 700 million people around the world. Deaf or mute people use hand gestures to communicate with others, to express themselves correctly. People who speak a natural language do not always understand their actions. To understand this, we need sign language interpreters. Their number is very small, and most of them work in large cities or regional centers. The solution to these problems can be found by human computer interaction. Today, a lot of research is being carried out in this direction. To solve this problem, a lot of research is currently underway. But the program that provides the perfect two-way translation has not yet been created. Especially for people who speak in Kazakh language. In this paper will be considered classical algorithms for gesture recognition.

Key words: Gesture recognition, KNN, Decision Tree, Naive Bayes, Logistic Regression, Classification.

1 Introduction

Communication between people is an integral part of life. In the process of communication, people convey their emotions, thoughts, desires to each other. People with disabilities, such as deaf and dumb people, experience various difficulties in the process of communication [1]. Today, 5% of the world's people (more than 430 million), and in Kazakhstan more than 18 thousand people suffer from deafness [2]. By 2050, this it is assumed will reach 700 million people around the world. Deaf or mute people use hand gestures to communicate with others, to express themselves correctly. People who speak a natural language do not always understand their actions. To understand this, we need sign language interpreters. Their number is very small, and most of them work in large cities or regional centers. The solution to these problems can be found by human computer interaction. Today, a lot of research is being carried out in this direction. To solve this problem, a lot of research is currently underway. But the program that provides the perfect two-way translation has not yet been created. Especially for people who speak in Kazakh language. In this paper will be considered classical algorithms for gesture recognition.

Various sensors are used to collect gestures data. The main sensors in use are optical sensors, inertial measurement units and radars. Inertial sensors provide more accurate data on hand movements and are independent of environmental conditions. But such sensors are inconvenient to use in certain areas due to the large number of wires. With the development of microwave technologies, radars are gaining popularity in the field of hand gesture recognition. In addition, radars will be independent of environmental conditions. But this method also has its own drawbacks, and it will be inconvenient to use radio wave encoders in everyday life. Due to the availability and cheapness of their devices, optical sensors are widely used today. Such sensors are also convenient in use and are used in many devices. For example, handheld smartphones, laptops, and similar devices used on a daily basis. But such sensors will depend on the ambient conditions, brightness [3][4][5].

Many methods are used to recognize static images. These include both classical methods and methods based on neural networks. However, classical methods differ in their simplicity and ease of understanding. Another advantage of these methods is that they are quick to learn.

2 Materials and Methods

Among these methods, the K-near-neighbors method, the naive Bayesian method, the decision

tree, and logistic regression methods are considered, which are widely used today.

The naive Bayes classifier is a simple probability classifier based on the use of Bayesian theorem with strict (naive) assumptions about independence.

Due to the specific nature of the probability model, naive Bayesian classifiers can be trained very effectively, and despite their simplicity, they can classify data well. Many practical applications for estimating parameters for naive Bayesian models use the maximum probability method, in other words, one can work with a naive Bayesian model without relying on Bayesian probabilities or using Bayesian methods [6].

A decision tree (also known as a classification tree or regression tree) is a decision — making tool used in machine learning, data analysis, and statistics. The structure of the tree is "leaves" and "branches". On the edges ("branches") of the decision tree, signs are written that depend on the target function, on the "leaves" the values of the target function are written, and on the rest of the nodes, signs are written in which the conditions differ. The goal of the decision tree is to build a model that predicts the value of the target variable using simple decision-making rules. Learning algorithms based on Decision Trees are considered one of the most widely used algorithms. The decision tree is used to solve classification and regression problems [7].

Logistic regression or Logit Model is a statistical model used to predict the probability of occurrence of some event by comparing it with a logistic curve. This regression gives a response in the form of the probability of a binary event (1 or 0). This method was introduced in 1958 by Dr. statistician Cox. This is a controlled method of machine learning used in classification problems [8].

The basic idea of logistic regression is to divide the space of initial values by a linear boundary (i.e. a straight line) into two regions according to classes. In the case of two dimensions, the linear boundary is simply a straight line without curves. This boundary is given depending on the available initial data and the learning algorithm. For the model to work correctly, the source data points must be divided by a linear boundary into the two areas mentioned above. If the initial data points satisfy this requirement, it can be said that they are linearly distributed.

The k – Nearest Neighbors method (kNN) is a popular classification algorithm used in various types of machine learning. Along with the decision tree, this is one of the most understandable ways to classify.

In the case of using the classification method, the object is assigned to the most common class among the K neighbors of this element, the classes of which will already be known.

In the case of using the regression method, the average value of k is assigned to the objects closest to the object, the values of which are already known [9].

3 Data collection

To collect hand gestures, we used the python programming language and the cvzone and cv2 modules of this programming language. For to capture images we used an RGB-type computer camera. The size of the images is 75x75x3. And in order to classify using the above methods, it is necessary to bring the data to the form of a one-dimensional vector. To do this, we use the gray format when reading from files. In this format, the images are two-dimensional. Then we can bring them to a one-dimensional form. After changing the shape, we get an array of size 5625x1.

To test the accuracy of machine learning methods, 5 types of hand gestures were used. They are the letters 'А', 'Ә', 'Б', 'В', 'Г'. Each letter contains 50 images. The total number of images is 250 for five letters.

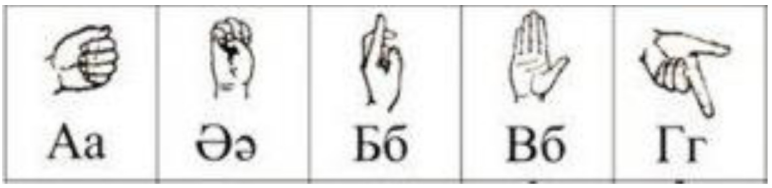

Figure 1 – Kazakh language letters in sign language

To classify images, we must first read them from a file and convert them to a one-dimensional array.

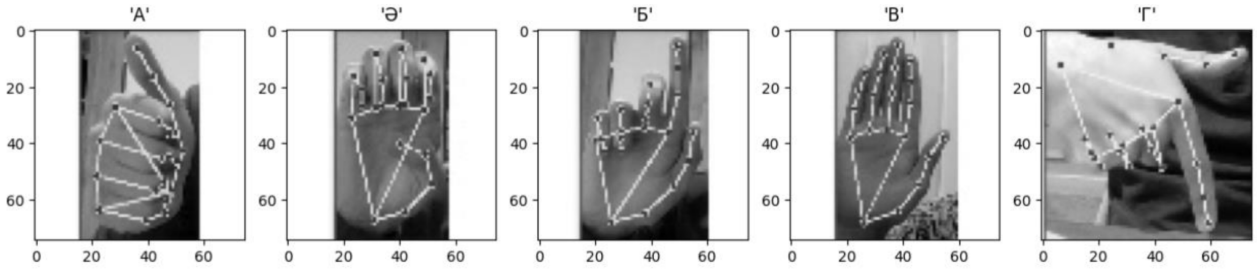

Figure 2 – Hand gestures on a gray map

	0	1	2	3	4	...	5621	5622	5623	5624	Class
0	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	1.000000	1.000000	A
1	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	1.000000	1.000000	A
2	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	1.000000	1.000000	A
3	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000	1.000000	1.000000	A
4	0.997195	1.000000	0.999435	0.998869	0.997195	...	1.000000	1.000000	1.000000	1.000000	A
...	...	...	...	...	...	...	...	...	...	...	...
245	0.999435	0.999435	1.000000	1.000000	0.997195	...	0.997195	1.000000	0.999167	0.997500	GG
246	0.997195	0.997195	1.000000	0.999435	0.999435	...	0.999435	0.997195	1.000000	1.000000	GG
247	0.997195	1.000000	0.995253	0.994687	0.996369	...	1.000000	1.000000	1.000000	1.000000	GG
248	0.998869	0.996629	0.994389	0.991584	0.997195	...	0.999435	0.996629	0.996629	0.997195	GG
249	0.870947	0.379605	0.448451	0.441732	0.452992	...	0.459567	0.369669	0.311411	0.269673	GG

250 rows × 5626 columns

Figure 3 – Images adjusted to one-dimensional type

After downloading the images and converting them to the desired format, we will analyze them for training and testing data. We will use 80% of the data for training and 20% for testing. Thus, 200 images will be trained, and 50 images will be tested.

We train and test models using the above models. Test result for each method:

Table 2 – The result of test

Algorithm	**Accuracy** (ratio of training data 80%)	**Accuracy** (ratio of training data 90%)
The naive Bayes classifier	0.72	0.6
Decision tree: criterion='entropy', splitter='best', max_depth=20, min_samples_split=10	0.86	0.92
Decision tree: criterion='entropy', splitter='best', max_depth=55, min_samples_split=10	0.88	0.92

Decision tree: criterion='gini', splitter='best', max_depth=20, min_samples_split=2	0.86	0.88
Decision tree: criterion='gini', splitter='best', max_depth=200, min_samples_split=2	0.8	0.88
Logistic regression penalty='l2', max_iter=70, solver='lbfgs', n_jobs=4	0.92	0.96
Logistic regression penalty='l2', max_iter=100, solver='saga', n_jobs=6	0.9	0.96
KNN n_neighbors=5, algorithm='auto', leaf_size=30, p=2, n_jobs=6	0.88	0.84
KNN n_neighbors=5, algorithm='ball_tree', leaf_size=50, p=1, n_jobs=6	0.94	0.96

4 Results

When 80% of all data was transmitted to the trained data, the accuracy of the models was less when compared with the 90%. Only in the Bayesian method was the accuracy less when the training data was provided with a 90%. The decision tree showed good results if criterion=entropy and min_samples_split=10. The Gini criterion is calculated faster than Entropy, but Entropy is more accurate [10]. And the logistic regression method has a good result when the amount of data is large (90%). The KNN algorithm showed good results when it passed the algorithm for calculating the nearest neighbors as 'ball_tree'. This is because the ball_tree algorithm calculates distances in big data faster than other algorithms.

The accuracy of other methods was higher than 90. The maximum accuracy was 96, which means that our model works with 4% error. This is a good result for classical machine learning methods.

5 Conclusion

In this work, images of hand gestures are collected and classical machine learning algorithms are implemented based on them. The best result was shown by the KNN algorithm and Logistic Regression with an accuracy of 96%. And we made sure that the correct allocation of data to training and testing data affects the accuracy of the algorithm. And as a result, we learned that the accuracy of algorithms depends directly on their parameters. If we give the right parameters, we will get high accuracy, even if we do not use a lot of data. In the future, we plan to continue this work with well-developed algorithms, such as neural networks and deep learning.

References

1. Y. Amirgaliyev, A. Aitimov , B. Amirgaliyev, B. Kynabay. “Comparative analysis of recognition algorithms for hand gestures on the basis of various representations of images”. Herald of the Kazakh-British technical university. № 1. (2019): 50-54.
2. Ayan KALMURAT. “V gluhonemyh vidyat lish' dvornikov i shvej”. February 14, 2019. https://rus.azattyq.org/a/kazakhstan-deaf-mutes-situation/29767359.html. (accessed 10.11.2022))
3. L. Wang, Z. Cui, Y. Pi, C. Cao және Z. Cao. “Low personality-sensitive feature learning for radar-based gesture recognition”. Neurocomputing. (2022): 373-384.
4. Weina Zhou, Kun Chen. “A lightweight hand gesture recognition in complex backgrounds”. Displays. (2022).
5. Fanchen Kong, Jingcheng Deng, Zichuan Fan. “Gesture recognition system based on ultrasonic FMCW and ConvLSTM model”. Measurement. (2022).
6. Pouria Kaviani, Sunita Dhotre. “Short Survey on Naive Bayes Algorithm”. International Journal of Advance Research in Computer Science and Management. № 11. (2019)
7. Sudha Bishnoi, B. K. Hooda. “Decision Tree Algorithms and their Applicability in Agriculture for Classification”. Journal of Experimental Agriculture International. № 7 (2022).
8. Prof. Ajit N.Gedam, Kajol B. Deshmane, Nishigandha N.Jadhav, Ritul M.Adhav, Akanksha N.Ghodake. “Survey on Breast Cancer Detection using Logistic Regression Algorithm”. International Journal of Innovative Research in Science, Engineering and Technology. № 4 (2022).
9. Shichao Zhang, Senior Member. “Challenges in KNN Classification”. Transactions on Knowledge and Data Engineering. (2021).
10. Samuel Aning, Malgorzata Przybyla-Kasperek. “Comparative Study of Twoing and Entropy Criterion for Decision Tree Classification of Dispersed Data”. Procedia Computer Science. (2022): 2434-2443