A Neural Network Classifier based American Sign Language Recognition System

Gauri Nagavkar[1] Kedar Potdar[2]
[1,2]Department of Electronics & Telecommunication Engineering
[1,2]Watumull Institute of Electronics Engineering and Computer Technology, Worli, Mumbai, India

Abstract— This paper introduces a communication system for the deaf-mute which helps them translate sign language into computer text using a hand-glove and neural network back-end. The hand-glove is fitted with flex sensors to measure finger movements. The electrical signals generated by the sensors are then processed and transmitted to a computer system with the use of an Arduino microcontroller. The processed signals are fed to a neural network trained to classify 26 letters of the alphabet. Use of neural networks for classification purposes has many significant advantages such as improved accuracy and flexibility of usage in different media. In this project, multiple combinations of feed-forward neural networks were used for the classification task. The network with 23 neurons outperformed rest of the networks with a high classification accuracy, evaluated with a low Mean Square Error of 0.510.

Key words: Neural Network Classifier, American Sign Language Recognition System

I. INTRODUCTION

Sign Language is the well-structured code gesture which uses hand gestures instead of sound to convey meaning, using orientations and movement of the hands. The Real Time Communication system for the dumb and deaf consists of an electronic device that can translate sign language into speech and text in order to make the communication possible between the deaf and/or mute with the general public. Gesture recognition is the mathematical interpretation of orientation or motion of human body by a computational system. This technology has been used in a variety of application areas, which demands accurate interpretation of sign language. [1]

The use of absolute values of the signals from flex sensors directly has several drawbacks. The system has to be reconfigured for each user since the degree of bending of fingers and the orientation of fingers may vary. Also, a slight change in the action of the same user might produce erroneous output. Thus, use of absolute signal values directly leads to less flexibility and accuracy. To overcome these limitations, the output is obtained from a neural network that has been trained to classify 26 letters of the English alphabet based on the normalized values of the electrical signals.

Section II describes the previous work that has been done in this field. American Sign Language has been introduced in Section III. Section IV describes Artificial Neural Networks and how they can be used for classification purposes. The system's architecture has been described in Section V. Section VI discusses the Neural Network approach used for ASL recognition. The system's working has been described in Section VII and Section VIII concludes our work.

II. RELATED WORK

About 70 million people in the world are speech and hearing impaired. There have been many projects till date which aim at bridging the communication gap between these especially abled people and their audience, using different techniques.

S. Upendran and Thamizharasi A. present a new method to recognize ASL alphabets from an image input. The system extracts Principal Component Analysis (PCA) features from the image of the ASL hand pose. The k-NN (k-Nearest Neighbour) classifier is used to distinguish between various gestures and their meanings. [2]

Anbarasi Rajamohan, Hemavathy R., Dhanalakshmi M developed Deaf-Mute Communication Interpreter which uses flex sensors and accelerometer to convert signs into nomal language. The text to speech conversion was done using TTS256 SpeakJet, which is a TTS module. Successful interpretation of only 10 letters (A, B, C, D, F, I, L, O, M, N, T, S, W) was achieved. [3]

In [4], the authors describe the use of sensory Cyberglove and a Flock of Birds 3-D motion tracker to extract the gestures and obtain an accuracy of 96% for the ASL alphabet.

In [5], the authors describe the structure and performance of the SLARTI sign language recognition system developed at the University of Tasmania. In this, CyberGlove was used to measure the degree of flexing and Polhemus IsoTrak was used to track the spatial position and orientation of the hand with respect to a fixed electro-magnetic source. These two input streams were processed and given as an input to a classifier network. An accuracy of around 94% was obtained on the signers (users) used in training, and about 85% for other signers.

[6] Describes a dynamic gesture recognition system for Korean Sign Language. Data acquisition is done a pair of Data-Gloves. A fuzzy min-max neural network is adopted for on-line pattern recognition to identify certain words from the Korean Sign Language. This system has an accuracy of 85% for the given set of words.

In [15], the authors have described an interpretation system for Indian Sign Language using video processing methods like e frame extraction, segmentation and refining of images, feature extraction and training a neural network to convert the video stream of words spoken using the Indian sign language into equivalent English sentences.

III. AMERICAN SIGN LANGUAGE

The American Sign Language is the sign language of the deaf people in the United States of America. Besides the USA, it is also used in several other regions like Canada, West Africa and South-East Asia. ASL is the second most widely used non-English language in the United States after Spanish. Like the English language, ASL has 26 letters and 6000 words.

Letters are expressed by using the right hand or the left hand or both hands, and facial expressions. Finger-spelling is the presentation of each letter of the alphabet by a sign. Each of the 26 letters has a different sign allocated to it. Finger-spelling is used to spell out Proper nouns and words that do not have any sign (words not belonging to the set of the 6000 words). [7]

IV. ARTIFICIAL NEURAL NETWORKS

Artificial Neural Networks (ANN) is a group of interconnected nodes that uses a computational model for information processing. It changes structure based on external or internal information that flows through the network. ANN can be used to model a complex relationship between inputs and outputs and find patterns in data. The output of ANN is determined by characteristics of the features and the weights associated with the interconnections among them. The connections between nodes are modified in the training process to adapt the network to desired outputs. [8]

The neural network gains the experience initially by training the system to correctly identify pre-selected examples of the problem. The response of the neural network is reviewed and the configuration of the system is refined until the neural network's analysis of the training data reaches a satisfactory level. In addition to the initial training period, the neural network also gains experience over time as it conducts analyses on data related to the problem. [9]

Classification using ANN is one of the most dynamic research and application areas. ANN is widely used for classification purposes because of its ability to generalize and map input-output relations based on existing data. [12]

V. SYSTEM ARCHITECTURE

This section describes the physical and logical architecture of the system. The system comprises of a hand glove fitted with flex sensors connected to an Arduino Uno microcontroller to detect finger movement as electrical signals. These signals are then parsed by a neural network running on a back-end computer to recognize the corresponding letter in ASL.

A. Tools Used

The components that make up the system are described below:

1) Flex Sensors

Flex sensors are variable resistors which change their resistance value when bent. When a flex sensor is bend, it changes resistance corresponding to its bend radius. Smaller the radius, higher is the resistance value. [10]

2) Arduino Uno

Arduino Uno is a microcontroller based on the ATmega328. It has 14 digital I/O pins, 6 analog inputs pins, a 16 MHz ceramic resonator, a USB connection, a power jack, an ICSP header and a reset button. It can be powered via the USB connection or with an external power supply (5 to 20 volts). [11]

3) Hand Glove

A hand glove, mounted with flex sensors for all the five fingers, is worn by the user. The hand glove must be necessarily thin, allowing for ease of movement of the sensors.

B. Hand Glove Assembly

Flex sensors are mounted on top of the hand glove as shown in Fig. 1. Each of the five sensors are then connected to power supply and analog input pins on the Arduino board. The analog input pins receive electrical signals to be further processed by the Arduino and backend system.

C. Arduino Uno Circuit

The Arduino Uno Circuit includes additional assembly to help relay the five flex sensors' values to the computer which are further processed there.

Two push buttons are used in the circuit. One for capturing the gesture, i.e. conversion of current gesture to corresponding letter from sign language, and one for printing the End-Of-Line character (EOL) through the Serial Port.

Fig. 1: Flex Sensors mounted on the Hand glove

The buttons are connected to two digital pins on the Arduino board.

D. Computer Backend

The Arduino microcontroller is connected to a computer running the neural network for classification via a USB port. A Backpropagation neural network is deployed on the computer which is trained to classify letters of the ASL based on signal inputs received from the Arduino Serial. [9]

VI. NEURAL NETWORK FOR LETTER CLASSIFICATION

To use a neural network for real-time classification tasks, it first needs to be trained on a labelled dataset and then deployed to perform classification based on live inputs. This section describes the procedure used to build a dataset, pre-processing it and training the network for classification.

A. Data Acquisition

A labelled dataset is needed for training the neural network to perform classification. Each letter in the alphabet corresponds to one position of the fingers. The Arduino Uno transmits five values to the computer, which acts as input to the neural network.

To develop an initial training dataset, 100 instances of sensor values were recorded for each letter, resulting in a dataset with 6 attributes and 2600 rows. The 6 attributes being 5 sensor values and one output value i.e. the letter corresponding to hand position.

The data attributes can be summarized as:

- R1 – thumb sensor value
- R2 – index finger sensor value
- R3 – middle finger sensor value
- R4 – ring finger sensor value
- R5 – little finger sensor values
- Output – character corresponding to current hand gesture

B. Preprocessing

The original data collected consists of five numerical attributes (sensor values) and one character attribute (output). The data needs to undergo process of treatment to before going to next process methods. [12]

1) Numeric Conversion

The output attribute is a character value which takes values from A to Z (inclusive). However, neural networks and other Machine Learning techniques cannot parse character and string values, hence they need to be converted to numeric values using an encoding technique.

Here, we have encoded the letters A to Z with their corresponding numeric position in the alphabet; with A being 1, B being 2 and so on.

2) Data Normalization

The sensor values R1 to R5 have different upper and lower bounds, which might negatively affect the performance of the neural network in terms of its classification accuracy.

Normalization refers to scaling of data between upper and lower bounds of the transfer functions of ANN.

The time series data was transformed into normalized data (in range 0 to 1) using the equation: [8]

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

C. Network Model Selection and Training

In this project, a Backpropagation neural network was used to classify letters in the ASL alphabet based on the flex sensor values.[14]

As discussed above, the dataset comprised of 2600 rows and 6 attributes. Of these 2600 rows, 70% (i.e. 1820 instances) were used for training the network, 15% (i.e. 390 instances) are used for network validation and 15% (390 instances) are used for network testing.

Various architectures of the network with different number of hidden layers were tested. The Levenberg-Marquardt training algorithm was used for training. [13]

Classification accuracy was measured in terms of the Mean Square Error (MSE). The best neural network model has the smallest MSE value. Once the smallest MSE is achieved the neural network stops training. This helps in extremely reliable and accurate prediction.

Table 1: Shows the MSE values for the different network architectures.

Number of Hidden Neurons	MSE
15	1.657
16	0.620
17	1.358
18	1.614
19	1.293
20	0.931
21	0.755
22	0.552
23	0.510
24	0.581
25	0.877

Table 1: MSE Values for different network architectures

As we can observe, the network with 23 hidden layers outperforms rest of the network architectures with a low MSE of 0.510. Hence, it is used for the classification task.

VII. Working

The working of the system can be broadly classified into two units as follows.

A. Sensor Data Retrieval and Transmission

The function of this unit is to read current sensor values and transmit the same to the computer backend. A voltage divider circuit is used between the Arduino Uno board and the sensors. It brings down the voltage range to lower values. The stepped down voltage is then supplied to the flex sensors mounted on each finger. The flex sensors act as variable resistors. As the sensor is flexed, the resistance across it increases, giving rise to different values of voltage for different orientations of the fingers. The Arduino Uno acts as ADC and converts these analog signals into corresponding digital signals. (These values are mapped into a higher range for the convenience of classification between the letters and their finger orientations.)

The two buttons connected to Arduino function are responsible for sending data to the computer. The 'fetch' button transmits the current sensor values to be processed by the neural network over Arduino's Serial output. The EOL button prints the End-of-Line character on Serial output.

B. Classification using neural networks

The sensor values transmitted to the computer are fed as inputs to the neural network. The network then classifies the input as one of the 26 letters.

The letter output can then be used as input for a display mechanism or a text-to-speech synthesis.

VIII. Conclusion

A neural network based classifier is developed to translate gestures in the ASL to text on a computer system, which can be further used for different purposes serving as another input mechanism for the deaf and dumb. [1]

References

[1] K. Potdar and G. Nagavkar, "Real-time Communication System for the Deaf and Dumb", International Journal of Innovative Research in Science, Engineering and Technology, Vol.6, Issue.4, pp.8338-45, 2017.

[2] S. Upendran, and A. Thamizharasi. "American Sign Language interpreter system for deaf and dumb individuals", In the Proceedings of the International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT), pp. 1477-1481, 2014.

[3] A. Rajamohan, R. Hemavathy and M. Dhanalakshmi "Deaf-Mute Communication Interpreter", International Journal of Scientific Engineering and Technology, Vol.2, Issue.5, pp.336-341, 2013.

[4] A. Oz and M.C. Leu, "Recognition of finger spelling of American sign language with artificial neural network using position/orientation sensors and data glove", Proceedings of the International Symposium on Neural Networks Springer, Berlin, Heidelberg, pp. 157-164, May 2005.

[5] P. Vamplew and A. Adams., "Recognition of sign language gestures using neural networks", Proceedings

of the European Conference on Disabilities, Virtual Reality and Associated Technologies, January 1996.
[6] J.S. Kim, W. Jang and Z. Bien, "A dynamic gesture recognition system for the Korean sign language (KSL)" IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 26, Issue.2, pp.354-359, 1996
[7] R.. Wilbur, "American Sign Language: Linguistics and Applied Dimensions", Little, Brown and Co, 1987
[8] K. Potdar and R. Kinnerkar, "A Comparative Study of Machine Learning Algorithms applied to Predictive Breast Cancer Data", International Journal of Science and Research, Vol.5, Issue.9, pp.1550-1553, 2016.
[9] K. Potdar and T. Pardawala, "Forecasting Ambient Air Quality in Mumbai using Neural Networks", Proceedings of the National Conference on Role of Engineers in Nation Building, Vol. 5, March 2017.
[10] F. Lotti, P. Tiezzi, G. Vassura, L. Biagiotti and C. Melchiorri. "UBH 3: an anthropomorphic hand with simplified endo-skeletal structure and soft continuous fingerpads", Proceedings IEEE International Conference on Robotics and Automation, 2004 (ICRA'04), Vol. 5, pp. 4736-474, IEEE, 2004.
[11] A. D'Ausilio, "Arduino: A low-cost multipurpose lab equipment" Behavior research methods, Vol. 44, Issue.2, pp.305-313, 2012.
[12] A. Lentz, "Machine Learning with R", Packt Publishing, 2005.
[13] J.J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory", Proceedings of Numerical analysis, Springer, Berlin, Heidelberg, pp. 105-115, 1978
[14] R. Hecht-Nielsen, "Theory of the backpropagation neural network. Neural Networks", 1(Supplement-1), pp.445-448, 1988.
[15] J. Rege, A. Naikdalal, K. Nagar and R. Karani, "Interpretation of Indian Sign Language through Video Streaming", International Journal of Computer Sciences and Engineering, Vol.3, Issue.11, pp.58-62, 2015.