Puzzle optimization algorithm with capsule networks based Telugu character recognition model

Josyula Siva Phaniram, Mukkamalla Babu Reddy
Department of Computer Science, Krishna University, Machilipatnam, India

ABSTRACT

Recently, Telugu character recognition (TCR) becomes a hot research topic because of drastic increase in technological advancements such as multimedia, and smartphones. Though numerous works have been concentrated on offline TCR models, it is still needed to develop automated and intelligent online TCR models. This paper presents a novel puzzle optimization algorithm with capsule networks-based Telugu character recognition (POACN-TCR) model. The presented POACN-TCR model intends to effectively identify and recognize distinct Telugu characters online. To accomplish this, the POACN-TCR model primarily undergoes pre-processing in different ways such as normalization, smoothing, and interpolation. In addition, the POACN-TCR model designs an effective capsule network (CapsNet) model to generate feature vectors and hyperparameter optimization takes place using puzzle optimization algorithm (POA). Finally, the C4.5 decision tree classifier is utilized for the effective recognition of Telugu characters. The utilization of POA for hyperparameter optimization of the CapsNet model helps in achieving improved recognition performance. For ensuring the enhanced outcomes of the POACN-TCR model, a wide-ranging experimental analysis is performed and the outcomes pointed out the betterment of the POACN-TCR model on the recent approaches.

Keywords: Data collection, Deep learning, Normalization, Puzzle optimization algorithm, pre-processing, Telugu character recognition, capsule network

This is an open access article under the CC BY-SA license.

Corresponding Author:
Josyula Siva Phaniram
Department of Computer Science, Krishna University
Machilipatnam, Andhra Pradesh, India
Email: phaniram.research@gmail.com

1. INTRODUCTION

Handwritten character detection was assumed that one of the appeals and stimulates research areas from the field of computer vision and pattern detection [1], [2]. Due to the major component of differences in cursive text and writing pattern, and the comparison of several characters from the procedure, recognition research was challenging and time utilized [3]. The online character detection was somewhat easy due to the temporal-based character properties such as procedure, count of distances, stroke, and writing direction. The offline character identification performance was complex due to their difference in fonts and writers [4], [5].

Telugu is the ancient and old language comprising Consonants-36, Vowels-16, and Guninthalu (Character generated by the combination of Telugu Vowels and Consonants) 560, around 612 characters [6]. Due to their multi classification problems, it could be difficult to recognize the handwritten Telugu characters, particularly, in Telugu Guninthalu, as all the characters are closely compared with one another. Many scientists are tried in identifying Indian languages particularly Telugu was explained from the survey to detect compound characters with template equivalent and several other approaches [7], [8]. Many machine learning (ML)
approaches were utilized to Pashto handwritten character recognition [9]. The power of deep neural networks (DNN) is for providing effectual implementation with no assuming the basic rule of deep element is linear [10].

Bhat et al. [11] present a method for the detection and classification of ancient Tigalari characters in the handwritten text. Tigalari is extremely utilized from coastal Karnataka and Kerala to document Malayalam, Sanskrit, and Tulu languages. This technique contains the design of database, proposal of deep convolutional neural network (DCNN) based framework for classifying the text, trained the method with data and identifying text utilizing test set. Prathima and Muppulaneni [12] refer to the handwritten Telugu vowels detection by utilizing a CNN method. The dataset is pre-processing and also feature extracting utilizing DNN method to train.

Jaiisswal et al. [13] generate a deep learning (DL) technique which is classifying up to 12 distinct languages, at document level and demonstrates the resultant dependence upon the outcome by executing line-level segmentation, then word-level segmentation, and lastly with character detection of the documents. Guha et al. [14] different CNN approach on publicly accessible hand-written Devanagari character and numeral dataset. It can mostly concentrate on comparative works by assuming memory utilization, training time, and trainable parameters. It can be planned and presented DevNet, an altered CNN structure that creates feasible outcomes since memory space and computation complexity are most important issues in this method [15]-[19].

This paper presents a novel puzzle optimization algorithm with capsule networks-based Telugu character recognition (POACN-TCR) model. The presented POACN-TCR model undergoes pre-processing in different ways such as normalization, smoothing, and interpolation. In addition, the POACN-TCR model designs an effective capsule network (CapsNet) model to generate feature vectors and hyperparameter optimization takes place using POA. Finally, the C4.5 decision tree classifier is utilized for the effectual recognition of Telugu characters. For ensuring the enhanced outcomes of the POACN-TCR model, a wide-ranging experimental analysis is performed.

2. THE PROPOSED MODEL

In this study, a novel POACN-TCR model has been developed to effectively identify and recognize distinct Telugu characters online. Followed by, the POACN-TCR model designing an effective CapsNet model to generate feature vectors and hyperparameter optimization takes place using POA. Finally, the C4.5 decision tree classifier is utilized for the effectual recognition of Telugu characters. Figure 1 illustrates the overall process of POACN-TCR technique.

![Figure 1. Overall process of POACN-TCR technique](image)

2.1. Data pre-processing

The POACN-TCR model primarily undergoes pre-processing in different ways such as normalization, smoothing, and interpolation. Data preprocessing, a component of data preparation, describes any type of processing performed on raw data to prepare it for another data processing procedure. It has traditionally been an important preliminary step for the data mining process. More recently, data preprocessing techniques have been adapted for training machine learning models and artificial intelligence (AI) models and for running inferences against them. Data preprocessing transforms the data into a format that is more easily and effectively processed in data mining, machine learning and other data science tasks. The techniques are generally used at the earliest stages of the machine learning and AI development pipeline to ensure accurate results.
Normalization: some characters are written by various strokes with any strokes expanded below or above an initial part of characters. An initial part of characters offers a measure of character size and line space employed by writers.

Smoothing: gaussian filter was executed to smoothen the stroke. A 1D Gaussian distribution was taken to account by the smoothing kernel which is primarily convolved with x- and y-axis. The smoothing was reached dependent upon the size of window functions and the smoothing kernel parameter utilized.

Interpolation: the final stage in pre-processed in which the smoothing stroke was included to offer a set amount of points, likewise spaced with curve length. The count of points was chosen based on the average amount of points to all the strokes from the provided data sets.

2.2. CapsNet feature extraction

Once data pre-processing is done, the next level is to extract feature vectors. CapsNets mentions a completely novel type of DL structure that attempts for conquering the limitation and disadvantages of CNN techniques as lose valued data and lack the precise model of entity at the time of max-pooling. The typical CapsNet is shallower, with 3 layers namely PrimaryCaps, DigitCaps, and Convld layers. The capsule-based representation of the group of hidden neurons, whereas probability and property of hidden features are captured. During this sample, the CapsNet is strong for affine conversion and minimum trained data. Furthermore, the CapsNet has resulted in particular developments connected to spatial hierarchies amongst features. The capsule signifies the group of neurons [20]. The activity of neurons encircled the different features of particular entity. While the length of resultant vectors signifies the probability of existence, the resultant capsule was computed by nonlinear squashing function.

\[ v_j = \frac{||s_j||^2 s_j}{\varepsilon + ||s_j||^2} \]

In which \( v_j \) denotes the vector outcome of capsules \( j \) and \( s_j \) refers the entire input. The non-linear squashing function is activation function to make sure that the short vector obtains shrunk to approximately zero length and longer vector obtains shrunk to particular length assuming \( \varepsilon \).

\[ s_j = \sum_i c_{ij} w_{ij} u_i \]

The entire inputs to capsule \( s_j \) was obtained by multiplying the outcome \( u_i \) of capsule utilizing a weighted matrix \( W_{ij} \) which signifies a weight amount on every predicted vector in the capsule in the under layer. At this point, \( c_{ij} \) signifies the coupling co-efficient for instance, determined as the iterative dynamic routing process.

\[ c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{jk})} \]

In which \( b_{ij} \& b_{jk} \) implies the log prior probabilities amongst 2 coupled capsules. The entire length of resultant vectors signifies the predicted chance. A separate margin loss \( L_k \) per capsule, \( k \) was:

\[ L_k = T_k \max (0, m^+ - ||v_k||) + \lambda (1 - T_k) \max (0, ||v_k|| - m^-)^2, \]

whereas \( T_k = 1 \), \( m^+ = 0.9 \), and \( m^- = 0.1 \) signifies the three free variables by default. \( \lambda \) allow the down weighted of loss and support for ensuring latter convergence.

2.3. Hyperparameter optimization

For optimally modifying the hyperparameters related to the CapsNet model, the puzzle optimization algorithm (POA) is applied. POA is presented and its arithmetical model is shown for resolving optimization problems [21]. A presented method is a population-based approach that is established according to puzzle game simulation. The population in POA is arithmetically modelled by a matrix that is shown in (5).

\[
X = \begin{bmatrix}
X_1 \\
\vdots \\
X_l \\
\vdots \\
X_N
\end{bmatrix}_{N \times m} = \begin{bmatrix}
X_{1,1} & \cdots & X_{1,d} & \cdots & X_{1,m} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
X_{l,1} & \cdots & X_{l,d} & \cdots & X_{l,m} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
X_{N,1} & \cdots & X_{N,d} & \cdots & X_{N,m}
\end{bmatrix}_{N \times m}
\]
Now, $X_i$ indicates the $i$'th puzzle, $X$ denotes the population of puzzles, $m$ represents the number of problem variables, $N$ represented the number of population of puzzles, and $x_{i,d,i}$ indicates the value of $d$'th variable recommended by $i$'th puzzle. Assume that every member of the population is a solution to the optimization issue, the value of objective function is estimated. Thus, equivalent to the number of population members, the objective function is estimated that the attained value for objective function is simulated.

$$F = \begin{bmatrix}
    f_1 \\
    f_2 \\
    \vdots \\
    f_{N^2}
\end{bmatrix} = \begin{bmatrix}
    F(X_1) \\
    F(X_2) \\
    \vdots \\
    F(X_N)
\end{bmatrix}_{N \times 1}$$

(6)

Now, $F$ denotes the vector of attained value for objective function and $f_i$ is represent value of objective function of $i$'th puzzle. As per the comparison of value attained for the objective function, the member that provides the optimal value for the objective function is known as the optimal member of the population. It is shown in:

$$B = X_k, f_k = \min (F)$$

(7)

where $B$ indicates the optimal member and $X_k$ denotes the $k$’th puzzle with minimal objective function equivalent to $f_k$. In the presented approach, population member is upgraded in two phases. Here, it is arithmetically expressed as:

$$GM_i = X_g, g \in \{1,2,3,...,N\}$$

(8)

$$dx_{i,d,i} = \begin{cases} 
    (GM_{i,d} - I \times x_{i,d,i}), & F_g < F_i \\
    (x_{i,d,i} - I \times GM_{i,d,i}), & \text{else}
\end{cases}$$

(9)

$$I = \text{round}(1 + \text{rand})$$

(10)

$$X_i^\text{new} = X_i + r \times dx_i$$

(11)

$$X_i = \begin{cases} 
    X_i^\text{new}, & f_i^\text{new} < F_i \\
    X_i, & \text{else}
\end{cases}$$

(12)

Now, $GM_{i,d}$ represent $d$’th dimension, $F_g$ denotes value of objective function,$GM_i$ denotes the guiding member of $i$’th puzzle, $dx_{i,d,i}$ signifies the change of $d$’th dimension of $i$’th puzzle, $I$ shows a random number that is 1 or 2, $r$ denotes an arbitrary value within $[0, 1]$ interval, $X_i^{\text{new}}$ indicates the novel status of $i$’th puzzle, and $f_i^{\text{new}}$ denotes the value of objective function. Next, every member of the population upgrades their status with the help of puzzle pieces recommended by other members of the population. It can be arithmetically expressed in (13)-(15).

$$N_p = \text{round} \left( 0.5 \times \left( 1 - \frac{1}{T} \right) \times N \right)$$

(13)

$$x_{i,d,j}^{\text{new}} = \begin{cases} 
    h \in \{1,2,3,...,N\} \\
    f \in \{1,2,3,...,N_p\} \\
    d_j \in \{1,2,3,...,m\}
\end{cases}$$

(14)

$$X_i = \begin{cases} 
    X_i^{\text{new}}, & f_i^{\text{new}} < F_i \\
    X_i, & \text{else}
\end{cases}$$

(15)

Where, $N_p$ denotes the number of recommended puzzle pieces, $T$ denotes the iteration counter, $T$ signifies the maximal amount of iterations, $X_{i,d,j}^{\text{new}}$ represent the new value for $d_j$’th dimension of $i$’th puzzle, and $x_{h,d,j}^{\text{new}}$ indicates the selected puzzle piece from $h$’th puzzle that $h$ is randomly chosen. Afterward upgrading each member of the population based on the initial and next phases, an iteration of the model, and the new status of the members of the population can be defined. Figure 2 illustrates the flowchart of POA.
2.4. Classification model

At the final stage, the C4.5 DT model is utilized for effective identification and recognition of Telugu characters. In the C4.5 DT approach experiments, the trained samples that are an identical outcome are eliminated as it is not most significant. Therefore, it is not be restricted in the DT once it does not have lesser outcomes that are minimum count of samples. The candidate splits are occupied as for concerning during the case that it is cut a specific count of samples. There is a multi district litigation (MDL) based adjustment for separating numeric attributes. Quinlan planned heuristic to avoid over-fit. The next subtraction is to determine that information gain (IG) is negative. Once it doesn’t have elements that are positive IG which is a variety of pre-pruning, the tree stops growing. It is signified as it can be unexpected to obtain a pruned tree but post-pruning is not active [22].

C4.5 employs the gain ratio:

\[ P(D; D_1, ..., D_k) = \frac{|D_k|}{D} \log \frac{|D_k|}{D} - \frac{1}{\log |D|} \sum_{i=1}^{k} \frac{|D_k|}{D} \log \frac{|D_k|}{D} \]  

(16)

which is distinct in the IG form, taking normalization on the count of feature values. Actually, the feature with maximal gain ratio, among features with higher-than-average IGs is selected as the separations.

3. EXPERIMENTAL VALIDATION

In this section, a detailed experimental validation of the POACN-TCR model is carried out using different aspects [23]. It is an integration of the 29,188 training instances gathered from 111 writers and 9,224 test instances gathered from 35 writers apart from the writers utilized for training data.

Table 1 provides a detailed performance validation of the POACN-TCR model under distinct aspects. Figure 5 shows the sample size and recognized samples offered by the POACN-TCR model. The results indicated that the POACN-TCR model has offered increased number of recognized samples. On top-1 results, the POACN-TCR model has recognized 4,513 samples out of original 4,690 samples. Also, on top-2
results, the POACN-TCR method has recognized 2,927 samples out of original 3,083 samples. Besides, on top-3 results, the POACN-TCR technique has recognized 950 samples out of original 997 samples. In addition, on top-4 results, the POACN-TCR model has recognized 174 samples out of original 191 samples. At last, on top-5 results, the POACN-TCR system has recognized 18 samples out of original 19 samples.

Table 1. Performances of proposed POACN-TCR model in terms of recognition rate (%)

| Methods      | Samples size | Recognized samples | Recognition rate (%) |
|--------------|--------------|--------------------|----------------------|
| POACN-TCR 1  | 4690         | 4513               | 96.23                |
| POACN-TCR 2  | 3083         | 2927               | 94.94                |
| POACN-TCR 3  | 997          | 950                | 95.29                |
| POACN-TCR 4  | 191          | 174                | 91.10                |
| POACN-TCR 5  | 19           | 18                 | 94.74                |
| No. of samples | 8980        | 8582               | 95.57                |

A detailed recognition rate examination of the POACN-TCR model under varying top-k results is given in Figure 6. The results indicated that the POACN-TCR model has showcased increased recognition rate. For instance, on top-1 results, the POACN-TCR model has offered recognition rate of 96.23%. At the same time, on top-3 results, the POACN-TCR algorithm has obtainable recognition rate of 95.29%. Meanwhile, on top-5 results, the POACN-TCR methodology has accessible recognition rate of 94.74%.

Finally, a comparative examination of the POACN-TCR model with recent models is provided in Table 2 and Figure 7 [24], [25]. The results indicated that the POACN-TCR model has demonstrated under distinct aspects. For instance, with top-1 results, the POACN-TCR model has offered increased recognition rate of 95.57% whereas the Babu et al. and Prasanth et al. models have provided reduced recognition rates of
91.60% and 90.60% respectively. Followed by, with top-3 results, the POACN-TCR approach has gained maximum recognition rate of 98.12% whereas the Babu et al. has accomplished minimal recognition rate of 97%. In addition, with top-5 results, the POACN-TCR approach has gained maximal recognition rate of 99.36% whereas Babu et al. [24] has accomplished reduced recognition rate of 98.70%. After examining the detailed results and discussion, it can be clear that the POACN-TCR model is found to be effective over the other methods.

| Performances | POACN-TCR | Babu et al. | Prasanth et al. |
|--------------|-----------|-------------|-----------------|
| Top-1        | 95.57     | 91.60       | 90.60           |
| Top-2        | 98.12     | 97.00       | -               |
| Top-3        | 98.78     | 98.00       | -               |
| Top-4        | 99.07     | 98.40       | -               |
| Top-5        | 99.36     | 98.70       | -               |

Table 2. Results comparison of proposed POACN-TCR method with existing models

Figure 7. Comparative analysis of POACN-TCR technique with existing methods

4. CONCLUSION

In this study, a novel POACN-TCR model has been developed to effectively identify and recognize distinct Telugu characters in online. The POACN-TCR model primarily undergoes pre-processing in different ways such as normalization, smoothing, and interpolation. Followed by, the POACN-TCR model designs an effective CapsNet model to generate feature vectors and hyperparameter optimization take place using POA. Finally, the C4.5 decision tree classifier is utilized for the effectual recognition of Telugu characters. The utilization of POA for hyperparameter optimization of the CapsNet model helps in achieving improved recognition performance. For ensuring the improved outcomes of the POACN-TCR approach, a wide-ranging experimental analysis is performed and the results pointed out the betterment of the POACN-TCR model on the recent approaches. In future, hybrid DL techniques are employed for enhanced TCR performance.

REFERENCES

[1] R. Aarthi, R. S. M. Varma, J. S. Vaishnavi, N. V. S. Prashanth, and G. T. Srikar, “Performance analysis of Telugu characters using deep learning networks,” in Lecture Notes in Electrical Engineering, vol. 711, 2021, pp. 311–322.
[2] N. Sarika and N. Sirisala, “Deep learning techniques for optical character recognition,” in Lecture Notes on Data Engineering and Communications Technologies, vol. 55, 2021, pp. 339–349, doi: 10.1007/978-981-15-8677-4_28.
[3] B. S. Babu, S. Nalajala, K. Sarada, V. M. Naidu, N. Yamnsi, and K. SaiKumar, “Machine learning based online handwritten telugu letters recognition for different domains,” in Intelligent Systems Reference Library, vol. 210, 2022, pp. 227–241.
[4] N. Sarika, N. Sirisala, and M. S. Velpuru, “CNN based Optical character recognition and applications,” in 2021 6th International Conference on Inventive Computation Technologies (ICICT), Jan. 2021, pp. 666–672, doi: 10.1109/ICICT50816.2021.9358735.
[5] K. R. Devarapalli and A. Negi, “Distributed training of deep neural network for segmentation-free Telugu word recognition,” in Machine Learning Technologies and Applications, Springer, Singapore, 2021, pp. 113–120.
[6] S. Inuganti, “Online handwritten Telugu character recognition using normalized differential chain code feature,” Webology, vol. 18, no. SI04, pp. 01–15, Sep. 2021, doi: 10.14704/WEB/V18S04/WEB19110.
[7] S. J. Basha, D. Veeraiah, G. Pavani, S. T. Afreen, P. Rajesh, and M. S. Sasank, “A novel approach for optical character recognition (OCR) of handwritten telugu alphabets using convolutional neural networks,” in 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICECS), Aug. 2021, pp. 1494–1500, doi: 10.1109/ICECS51422.2021.9532638.
[8] S. K. Gorla, S. S. Tangeda, L. B. M. Neti, and A. Malapati, “Telugu named entity recognition using bert,” International Journal of Data Science and Analytics, Jan. 2022, doi: 10.1007/s41060-021-00305-w.
[9] A. T. Anju, B. P. Chacko, and K. P. M. Basheer, “Review of offline handwritten text recognition in south Indian languages,” Malaya Journal of Matematik, vol. 9, no. 1, pp. 751–756, 2021, doi: 10.26637/MJM0901/0132.
Josyula Siva Phaniram has completed Master of Science in Computer Science from Nagarjuna University and Master of Technology in Information Technology from Karnata University, Machilipatnam. His research interest includes artificial intelligence, machine learning, deep learning. He can be contacted at email: phaniram.research@gmail.com.

Dr. Mukkamalla Babu Reddy has completed his Post Graduation and Doctoral degree from Acharya Nagarjuna University, AP. He is activity involved in teaching and research for more than two decades. He has published more than 60 research papers in reputed journals and guiding PhD Scholars in Computer Science & Engineering domain. At present, he is working as Principal, University College of Engineering and Technology, Krishna University, Machilipatnam, AP, India. He can be contacted at email: m_babureddy@yahoo.com.