International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

NLP Based Mathematical Descriptor

Prof. Sharada H N (Guide)
Dept of Computer Science and Engineering, SDM College of Engineering and Technology, Dharwad.

Prof. Sandya S V
Dept of Computer Science and Engineering, SDM College of Engineering and Technology, Dharwad

Prajwal M C
Student(UG), Dept of Computer Science and Engineering SDM College of Engineering and Technology, Dharwad

Pujari Vamshi
Student(UG), Dept of Computer Science and Engineering SDM College of Engineering and Technology, Dharwad

T Vishwanath Setty
Student(UG), Dept of Computer Science and Engineering SDM College of Engineering and Technology, Dharwad

Sanket Shetti
Student(UG), Dept of Computer Science and Engineering SDM College of Engineering and Technology, Dharwad

Abstract: ***Reading of mathematical expression or equation in the document images is very challenging due to the large variability of mathematical symbols and expressions. Here, we pose reading of mathematical equation as a task of generation of the textual description which interprets the internal meaning of this equation. Inspired by the natural image captioning problem in computer vision, we present a mathematical equation description (MED) model, a novel end-to-end trainable deep neural network-based approach that learns to generate a textual description for reading mathematical equation images.***

Our MED model consists of a convolution neural network as an encoder that extracts features of input mathematical equation images and a recurrent neural network with attention mechanism which generates description related to the input mathematical equation images.

I. Introduction

([1] Ajoy Mondal and C V Jawahar " Texual Description for Mathematics Equations", center for Visual Information Technology. 2019,
We have referred the above mentioned paper for implementing our project. In our implementation we have created the dataset but in the referred paper they took printed dataset and we also implemented the Voice output of the Equation Description.)

Our mathematical equation description model aims to generate textual descriptions that interpret the internal meaning of mathematical equations read from document images. The model comprises a convolutional neural network (CNN) serving as an encoder and a recurrent neural network (RNN) with an attention mechanism responsible for generating descriptions related to the input equation images.

To overcome the lack of a mathematical equation image dataset with accompanying textual descriptions, we created our own dataset for experimental purposes. This dataset allowed us to assess the effectiveness of our mathematical equation description model in reading and interpreting equations accurately.
Most existing automatic math reading systems primarily accept equations in the form of image or similar markup languages. In contrast, our approach focuses on directly processing equation images, presenting a unique challenge that requires specialized techniques.

The CNN encoder plays a crucial role in the model by extracting relevant features from the input equation images. By employing convolutional layers, the model can capture local patterns and structures, enabling it to learn meaningful representations of the equations. These extracted features are then passed to the RNN component.

The RNN component, equipped with an attention mechanism, generates natural language descriptions that pertain to the input mathematical equations. The attention mechanism allows the model to focus on different regions of the equation images, ensuring that the generated descriptions are contextually relevant and accurately convey the internal meaning of the equations.

II. Literature Survey

Automatic recognition of mathematical expressions (MEs) plays a critical role in transcribing scientific and engineering documents into digital form. This task involves two primary steps: symbol recognition and structural analysis. Symbol recognition begins with segmenting the symbols and then identifying each segmented symbol. The recognized symbols are then subjected to structural analysis to determine the mathematical expressions. These problems can be tackled sequentially or within a single global framework.[1]

However, both sequential and global approaches have their limitations. Segmentation of mathematical symbols poses challenges in both printed and handwritten documents due to the presence of a mixture of text, expressions, and figures.[2]

Symbol recognition is difficult due to the vast number of symbols, fonts, typefaces, and font sizes that need to be accounted for. For structural analysis, a commonly used method is employing two-dimensional context-free grammars, which require prior knowledge for defining math grammar. Additionally, the complexity of the parsing algorithm increases with the size of the math grammar.[3]

To address these challenges, models have been developed that utilize multi-layer convolutional networks to extract features from images. These models incorporate attention-based recurrent neural networks as decoders, enabling them to generate structured markup text. In a similar vein, a novel end-to-end approach based on neural networks has been proposed. This approach focuses on learning to recognize handwritten mathematical expressions (HME) in a two-dimensional layout and produces output as a one-dimensional character sequence.[4]

By leveraging these models and approaches, the automatic recognition of mathematical expressions can be significantly enhanced. These advancements have the potential to improve the efficiency of transcribing scientific and engineering documents into a digital format, facilitating further analysis and processing in these fields.[5]

While challenges exist in symbol segmentation, symbol recognition, and structural analysis, the integration of convolutional and recurrent neural networks has shown promise in addressing these limitations. The development of end-to-end approaches further streamlines the recognition process. As research in this area continues to evolve, the accuracy and reliability of ME recognition systems will continue to improve, benefiting various applications in the scientific and engineering communities.[7]

Understanding Human Language Those questions are interesting, but out of scope for this course. Those questions are the ones linguists try to answer. Generative linguistics aims at figuring out what those rules are, how they are combined to form a valid sentence, how they are adapted to different languages and so on. We will leave these to linguists and continue on to our journey of building a machine that understands human languages.[12]

Deep convolutional neural networks have led to a series of breakthroughs for image classification .Deep networks naturally integrate low/mid/highlevel features and classifiers in an end-to-end multilayer fashion, and the “levels” of features can be enriched by the number of stacked layers (depth). Recent evidence reveals that network depth is of crucial importance, and the leading results on the challenging ImageNet dataset all exploit models, with a depth of sixteen to thirty [10].

III. Detailed Design

A **Symbol segmenter** is a type of natural language processing (NLP) model that is used to identify and segment symbols in a text. One common approach is to use a rule-based system. Rule-based systems use a set of hand-crafted rules to identify and segment symbols.

A **Symbol classifier** is a machine learning model that identifies and categorizes symbols in images or text, including numbers, letters, punctuation marks, and special characters. It is used in applications like machine translation, text analysis, and information retrieval to extract specific symbols from data.

A **Relation classifier** is a machine learning model that is trained to identify the relationship between two entities in a sentence. One common approach is to use a supervised learning model, which is trained on a dataset of sentences that have been manually labeled with the correct relationship. Another approach is to use an unsupervised learning model, which is trained on a dataset of sentences without any labels.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

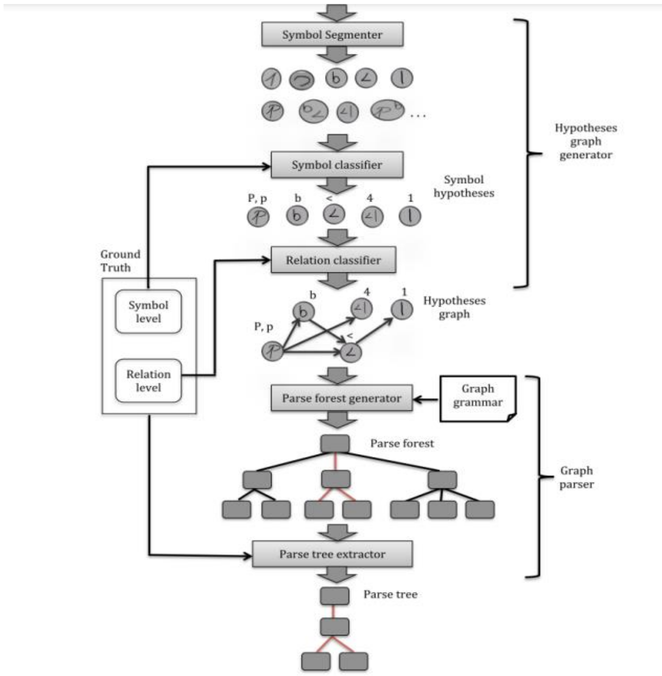


Fig.(a) Architecture Diagram

A **Parse forest generator** takes a grammar as input and automatically generates source code that can parse streams of characters using the grammar. The generated code is a parser, which takes a sequence of characters and tries to match the sequence against the grammar.

A **Parse tree extractor** is a program that takes a sentence as input and produces a parse tree as output. A parse tree is a tree-like structure that represents the syntactic structure of a sentence. The root of the parse tree represents the entire sentence, and the children of each node represent the constituents of the sentence.

Below model uses an end-to-end trainable network consisting of CNN followed by a language generation. It generates textual description of an input mathematical expression image in natural language which interprets its internal meaning.

IV. Implementation

The training dataset is the portion of the dataset that the model is exposed to during the training phase. It is used to train the model by adjusting its parameters and finding the best possible configuration for the model to fit the data. The training dataset typically consists of a large number of labelled examples that are used to teach the model how to make accurate predictions.

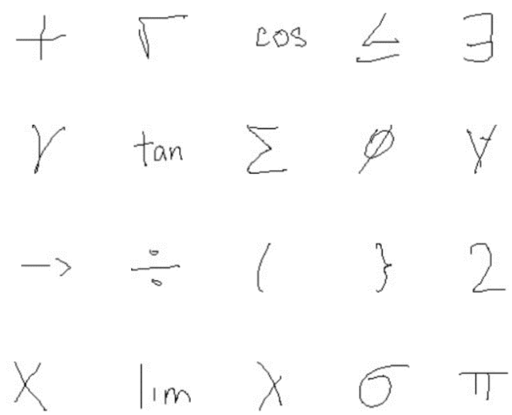
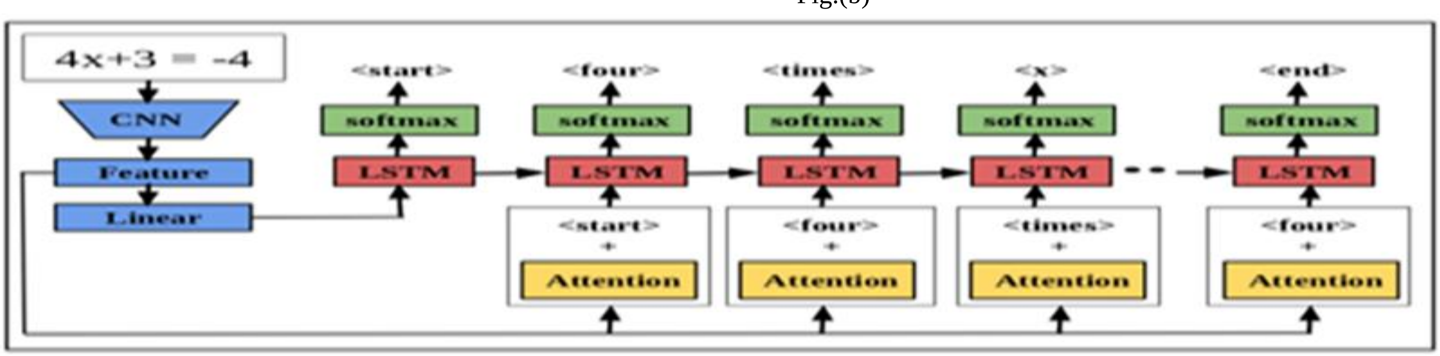


Fig. Dataset

The testing dataset is a portion of the dataset that is used to evaluate the performance of the trained model. It is a separate set of labelled examples that the model has not seen during the training phase. The testing dataset is used to measure how well the model generalizes to new, unseen data. If the model performs well on the testing dataset, it suggests that the model

Fig.(b)



International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

has learned to generalize from the training dataset and can make accurate predictions on new data.

It is important to keep the training and testing datasets separate to avoid overfitting, which occurs when the model learns to fit the training data too well and performs poorly on new, unseen data. To avoid this, a common practice is to split the dataset into training and testing datasets randomly. A

Work Flow

typical split is to use 70-80% of the data for training and the remaining 20-30% for testing.

Pre-processing: Mathematical equation descriptions undergo pre-processing using basic tokenization. Only words appearing at least four times in the training set are retained. This filtering steps ensures that the model focuses on frequently occurring and relevant words, improving its understanding and processing capabilities.

Training Details: The training objective of our Mathematical Equation Description model is to maximize the predicted word probability. We use cross entropy as the objective function.
Decoding: In decoding stage, our main aim is to generate a most likely textual description for a given Mathematical Equation Interpreter. classification, text classification, and speech recognition, where the input data has a sequential structure. The layers of the model can be configured with different types of operations, such as convolution, pooling, activation, and dropout layers.
As per our knowledge goes, no mathematical expression data sets with their textual descriptions are available for experiment. We create a data set, with large number of synthetically generated MEIs and their descriptions. For this purpose, we create sets of predefined functions (e.g., linear equation, limit, etc.), variables (e.g.: x, y, z, etc.), operators (e.g., +, −, etc.) and constants (e.g., 10, 1, etc.) and sets of their corresponding textual descriptions.

We develop a python code which randomly selects a function, variable, operator and constant from the corresponding predefined sets and automatically generates mathematical equation as an image and corresponding textual description in the text format.

V. Conclusion

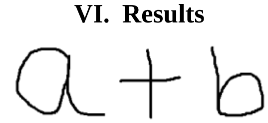
Our research introduces an innovative Mathematical Equation Description (MED) model aimed at aiding blind and visually impaired students in reading mathematical equations. The key aspect of our approach is the generation of textual descriptions for these equations. Unfortunately, the unavailability of mathematical images and their associated textual descriptions prompted us to create two distinct datasets specifically designed for experimentation.

To evaluate the practical application of our MED model, we conducted real-world experiments involving blind and visually impaired students. The results of these experiments were encouraging, as they demonstrated that the students were able to effectively write mathematical expressions after reading or listening to the descriptions generated by the MED network.

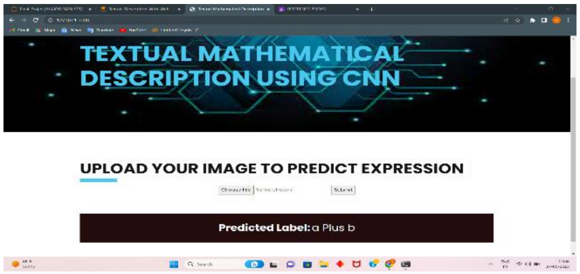
This successful outcome strongly establishes the effectiveness of our MED framework in facilitating the understanding and utilization of mathematical equations by students with visual impairments. By providing accurate and comprehensive textual descriptions, our model empowers these students to comprehend and work with mathematical expressions, fostering their engagement and learning in the subject matter.

The implications of our research are significant, as it addresses a crucial challenge faced by blind and visually impaired individuals in accessing and comprehending mathematical content. By leveraging the power of machine learning and natural language generation, our MED model opens up new avenues for inclusive education and empowers visually impaired students to actively participate in mathematics-related studies.

VI. Results



The input is given as above in image formate and output will be generated as below.



VII. References

[1] Ajoy Mondal and C V Jawahar " Texual Description for Mathematics Equations", center for Visual Information Technology. 2019

[2] F. Alvaro, J.-A. Sanchez, and J.-M. Benedf , "An integrated grammar-based approach for mathematical expression recognition," PR, 2016.

[3] W. Wongkia, K. Naruedomkul, and N. Cercone, "i-math: Automatic math reader for students," Computers & Mathematics with Applications, 2012.

[4] N. Soiffer, "MathPlayer: web-based math accessibility," in International ACM SIGACCESS Conference on Computers and Accessibility, 2005.

[5] J. Zhang, J. Du, S. Zhang, D. Liu, Y. Hu, J. Hu, S. Wei, and L. Dai, "Watch, attend and parse: An end-to-end neural network-based approach to handwritten mathematical expression recognition," PR, 2017.

[6] F. Julca-Aguilar, H. Mouchere, C. Viard-Gaudin, and N. S. `Hirata, "Top-down online handwritten mathematical expression parsing with graph grammar," in IberoAmerican Congress on PR, 2015.

[7] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in CVPR, 2015.

[8] Q. You, H. Jin, Z. Wang, C. Fang, and J. Luo, "Image captioning with semantic attention," in CVPR, 2016.

[9] K. Xu, J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," in ICML, 2015.

[10] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[11] K. Cho, "Natural language understanding with distributed representation," arXiv preprint arXiv:1511.07916, 2015.

[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in CVPR, 2009.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in CVPR, 2016.