ISSN: 0974-5823 Vol. 6 No. 3 December, 2021

International Journal of Mechanical Engineering

DESIGN AND ANALYSIS OF ADVANCED APPRAOCH FOR TEXT RECOGNITION USING CONVOLUTIONAL NEURAL NETWORKS

Research Scholar - **SARALADEVI ANDHE**[1]

Department of COMPUTER SCIENCE & ENGINEERING, School of Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore,MP,India

Research Guide - **Dr.Neeraj Sharma**[2]

Department of COMPUTER SCIENCE & ENGINEERING, School of Engineering , Sri Satya Sai University of Technology & Medical Sciences, Sehore,MP,India

Research Co-Guide - **Dr. B.Kavitha Rani** [3]

Department of COMPUTER SCIENCE & ENGINEERING, CMR Technical Campus, Hyderabad, Hyderabad, Telangana 501401.

Abstract-

In the area of pattern recognition and pattern matching, the methods based on deep learning models have recently attracted several researchers by achieving magnificent performance. In this research, we propose the use of the convolutional neural network to recognize the handwritten characters in an unconstrained environment. We also propose a novel dataset of handwritten characters since there is no publicly-available dataset of this kind. A series of experiments are performed on our proposed dataset. The accuracy achieved for character recognition is among the best while comparing with the ones reported in the literature for the same task. OCR stands for Optical Character Recognition and is the mechanical or electronic translation of images consisting of text into the editable text. It is mostly used to convert handwritten (taken by scanner or by other means) into text. Human beings recognize many objects in this manner our eyes are the "optical mechanism." But while the brain "sees" the input, the ability to comprehend these signals varies in each person according to many factors. Digitization of text documents is often combined with the process of optical character recognition (OCR). Recognizing a character is a normal and easy work for human beings, but to make a machine or electronic device that does character recognition is a difficult task. Recognizing characters is one of those things which humans do better than the computer and other electronic devices. With its astronomical growth over the past decade, the Web becomes huge, diverse and dynamic. Data mining is extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations. The application of data mining techniques to the web is called Web Mining. Web Mining aims to discover interesting patterns in the structure, the contents and the usage of web sites.

Keywords: Information Extraction, Handwriting Recognition; Convolutional Neural Network

1. INTRODUCTION

In the field of pattern recognition and computer vision research, the task of handwritten text recognition is regarded as one of the most challenging areas. The cursive nature of text, the shape similarity of individual characters, and the availability of different writing styles are some of the key issues that make the recognition task more challenging. While recognizing the isolated word and character in the printed text, higher accuracy rates are observed in the literature; however, there is a need for an efficient recognition system that gives remarkable results in recognizing handwritten texts [1].

1.1 BACKGROUND

In character recognition (also optical character reader, character recognition) the of images of handwritten, typed, or printed text is converted into machine-encoded text by the means of mechanical or electronic conversion, whether from a scanned document, an image of a document, a scene-image (for example the text on signs and billboards in a landscape image) or from subtitle text superimposed on an image (for example from a television broadcast) [2]. It is generally utilized as a type of data section from printed paper information records, regardless of whether international ID reports, solicitations, bank articulations, automated receipts, business cards, mail, printouts of static-information, or any appropriate documentation [3] It is a typical technique for describing printed messages with the goal that they can be electronically altered, looked, put away more minimally, showed on-line, and utilized as a part of machine procedures, for example, intellectual figuring, machine interpretation, (extricated) content to-discourse, key information and content mining. Character acknowledgment is a field of research in design acknowledgment; counterfeit consciousness and PC vision [4].

1.2 CONVOLUTIONAL NEURAL NETWORK

CNN has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision [5]. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm [6].

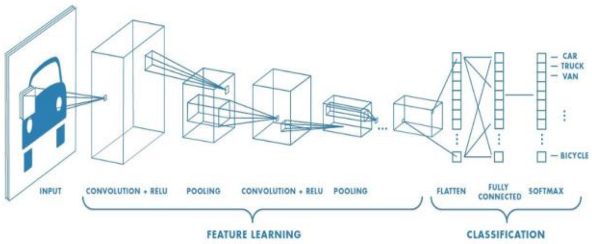

Figure 1.1 Convolutional Neural Network working Flow

1.3 MOTIVATION

Current ways to deal with protest acknowledgment make fundamental utilization of machine learning strategies. To move forward their execution, we can gather bigger datasets, take in more effective models, and utilize better methods for anticipating over fitting. Up to this point, datasets of named pictures were moderately little — on the request of a huge number of pictures (e.g., NORB , Caltech-101/256 [8, 9], andCIFAR-10/100 [10]). Basic acknowledgment undertakings can be fathomed great with datasets of this size, specifically on the off chance that they are increased with mark protecting changes. For instance, the current best mistake rate on the MNIST character-acknowledgment errand (<0.3%) approaches human execution .But protests in reasonable settings show extensive changeability, so to figure out how to remember them it is important to utilize significantly bigger preparing sets.

1.4 OBJECTIVES

The objectives of the present work include the following:

1. To design and develop recognize character creation model which is function based on traversing each and every pixel image using CNN.
2. To train the system to accurately identify characters use of TensorFlow. Model come across to thousands of examples of images to train itself using CNN.

3. Train and enhance the system using convolutional neural network and fully connected layers Fully connected layers is used rather than CNN to reduce the complexion ,getting high accuracy in less time.
4. To check the accuracy of the system with our testing data.

1.5 PROBLEM STATEMENT

The data stored in the web spaces are numerous and one can refer to any kind of information with the help of websites. Recently, information extracted from the web using programmed methods because of the need of information. As the extraction process becomes viral, the websites are becoming sources of redundant information. Duplication becomes a major issue. Thus, a method needed to extract information from the websites by identifying the relevant information. The main problem faced by extractors is that, a single website contains the same content a number of times and possesses other irrelevant information.

2. REVIEW OF LITERATURE

M.S.B. PhridviRaja, and C.V. GuruRaob (2014) have presented Data Stream Mining is one of the area gaining lot of practical significance and is progressing at a brisk pace with new methods, methodologies and findings in various applications related to medicine, computer science, bioinformatics and stock market prediction, weather forecast, text, audio and video processing to name a few. Data happens to be the key concern in data mining. With the huge online data generated from several sensors, Internet Relay Chats, Twitter, Face book, Online Bank or ATM Transactions, the concept of dynamically, changing data is becoming a key challenge, what we call as data streams. In this paper, they proposed algorithm for finding frequent patterns from data streams with a case study and identify the research issues in handling data streams.

Bing Liu, et al (2003) have developed a large amount of information on the Web is contained in regularly structured objects, which he call data records. Such data records are important because they often present the essential information of their host pages, e.g., lists of products or services. It is useful to mine such data records in order to extract information from them to provide value-added services. Existing automatic techniques are not satisfactory because of their poor accuracies. In this paper, he proposed a more effective technique to perform the task. The technique based on two observations about data records on the Web and a string-matching algorithm. The proposed technique is able to mine both contiguous and noncontiguous data records. Our experimental results show that the proposed technique outperforms existing techniques substantially.

Qiang Yang, Xindong WU (2006) have discussed, In October 2005, he took an initiative to identify 10 challenging problems in data mining research, by consulting some of the most active researchers in data mining and machine learning for their opinions on what are considered important and worthy topics for future research in data mining. He hope their insights will inspire new research efforts, and give young researchers (including PhD students) a high-level guideline as to where the hot problems are located in data mining. Due to the limited amount of time, he were only able to send out our survey requests to the organizers of the IEEE ICDM and ACM KDD conferences, and we received an overwhelming response. He was very grateful for the contributions provided by these researchers despite their busy schedules.

Soumya Ray and Mark Craven (2001) have studied the application of Hidden Markov Models (HMMs) to learning information extractors for n-ary relations from free text. He proposed an approach to representing the grammatical structure of sentences in the states of the model. We also investigate using an objective function during HMM training which maximizes the ability of the learned models to identify the phrases of interest. We evaluate our methods by deriving extractors for two binary relations in biomedical domains. His experiments indicate that our approach learns more accurate models than several baseline approaches.

Neelamadhab Padhy, et al (2012) have focused a variety of techniques, approaches and different areas of the research which are helpful and marked as the important field of data mining Technologies. As he was, aware that many MNC’s and large organizations operated in different places of the different countries. Each place of operation may generate large volumes of data. Corporate decision makers require access from all such sources and take strategic decisions. The data warehouse used in the significant business value improved the effectiveness of managerial decision-making. In an uncertain and highly competitive business environment, the value of strategic information systems such as these easily recognized however in today's business environment, efficiency, or speed is not the only key for competitiveness. These types of huge amount of data's are available in the form of tear- to peta-bytes, which has drastically changed in the areas of science and engineering.

Barahate Sachin R. and Shelake Vijay M. (2012) have proposed techniques for exploring, and analyzing the huge data that come from the educational context. EDM is poised to leverage an enormous amount of research from data mining community and apply that research to educational problems in learning, cognition, and assessment. In recent years, Educational data mining has proven to be more successful at many of these educational statistics problems due to enormous computing power and data mining algorithms. This paper surveys the history and applications of data mining techniques in the educational field. The objective is to introduce data mining to traditional educational system, web-based educational system, intelligent tutoring system, and e learning. This paper describes how to apply the main data mining techniques such as prediction, classification, relationship mining, clustering, and social area networking to educational data.

Edin Osmanbegovic, and Mirza Suljic (2012) have presented data mining has been successfully implemented in the business world for some time now, its use in higher education is still relatively new, i.e. its use is intended for identification and extraction of new and potentially valuable knowledge from the data. Using data mining the aim was to develop a model, which can derive the conclusion on students' academic success. Different methods and techniques of data mining were compared during the prediction of students' success, applying the data collected from the surveys conducted during the summer semester at the University of Tuzla, the Faculty of Economics, academic year 2010-2011, among first year students and the data taken during the enrollment. The success evaluated with the passing grade at the exam. The impacts of students' socio-demographic variables achieved results from high school and from the entrance exam, and attitudes towards studying, which can have an effect on success, were investigate. In future investigations, with identifying and evaluating variables associated with process of studying, and with the sample increase, it would be possible to produce a model, which would stand as a foundation for the development of decision support system in higher education.

Farkhund Iqbal, et al (2011) have presented the cyber world provides an anonymous environment for criminals to conduct malicious activities such as spamming, sending ransom e-mails, and spreading botnet malware. Often, these activities involve textual communication between a criminal and a victim, or between criminals themselves. The forensic analysis of online textual documents for addressing the anonymity problem called authorship analysis is the focus of most cybercrime investigations. Authorship analysis is the statistical study of linguistic and computational characteristics of the written documents of individuals. This paper is the first work that presents a unified data mining solution to address authorship analysis problems based on the concept of frequent pattern-based write print. Extensive experiments on real-life data suggest that our proposed solution can precisely capture the writing styles of individuals.

Cristobal Romero and Sebastian Ventura (2013) have presented applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from educational environments. Its goal is understand how students learn and identify the settings in which they learn to improve educational outcomes and to gain insights into and explain educational phenomena.

Nicholas Kushmerick (2000) has presented numerous sources of useful information telephone directories, product catalogs, stock quotes, event listings, etc. Recently, many systems have built that automatically gather and manipulate such information on a user's behalf. However, these resources are usually formatted for use by people (e.g., the relevant content is embedded in HTML pages), so extracting their content is difficult. Most systems use customized wrapper procedures to perform this extraction task. Unfortunately, writing wrappers is tedious and error-prone. As an alternative, we advocate wrapper induction, a technique for automatically constructing wrappers. In this article, we describe six wrapper classes, and use a combination of empirical and analytical techniques to evaluate the computational tradeoffs among them. He measured the number of examples and time required to learn wrappers in each class, and compare these results to PAC models of our task and asymptotic complexity analyses of our algorithms. Summarizing our results, he find that most of our wrapper classes are reasonably useful (70% of surveyed sites can be handled in total), yet can rapidly learned (learning usually requires just a handful of examples and a fraction of a CPU second per example).

Bing Liu, et al (2003) described latent Dirichlet allocation (LDA), a generative probabilistic model for collections of discrete data such as text corpora. LDA is a three level hierarchical Bayesian model, in which each item of a collection is model as a finite mixture over an underlying set of topics. Each topic is in turn, modeled as an infinite mixture over an underlying set of topic probabilities. In the context of text modeling, the topic probabilities provide an explicit representation of a document. We present efficient approximate inference techniques based on variation methods and an EM algorithm for empirical Bayes parameter estimation. He reported results in document modeling, text classification, and collaborative filtering, comparing to a mixture of unigrams model and the probabilistic LSI model.

Tak-Lam Wong and Wai Lam (2004) have developed a probabilistic framework for adapting information extraction wrappers with new attribute discovery. Wrapper adaptation aims at automatically adapting a previously learned wrapper from the source Web site to a new unseen site for information extraction. One unique characteristic of our framework is that it can discover new or previously unseen attributes as well as headers from the new site. It based on a generative model for the generation of text fragments related to attribute items and formatting data in a Web page. To solve the wrapper adaptation problem, he considered two kinds of information from the source Web site. The first kind of information is the extraction knowledge contained in the previously learned wrapper from the source Web site. The second kind of information is the previously extracted or collected items. His employ a Bayesian learning approach to automatically select a set of training examples for adapting a wrapper for the new unseen site. To solve the new attribute discovery problem, we develop a model, which analyzes the surrounding text fragments of the attributes in the new unseen site.

Paul Viola (2005) have proposed, conditional Markov chain models (CMM) have been used to extract information from semi-structured text (one example is the Conditional Random Field). Applications range from finding the author and title in research papers to finding the phone number and street address in a web page. The CMM framework combines a priori knowledge encoded as features with a set of labeled training data to learn an efficient extraction process. He will show that similar problems solved more effectively by learning a discriminative context free grammar from training data. The grammar has several distinct advantages: long range, even global, constraints used to disambiguate entity labels; training data is used more efficiently; and a set of new more powerful features can be introduced. The specific problem he considers is of extracting personal contact, or address, information from unstructured sources such as documents and emails. While linear-chain CMMs perform reasonably well on this task, he shows that a statistical parsing approach results in a 50% reduction in error rate. In cases where there are multiple errors, a single user correction propagated to correct multiple errors automatically. Using a discriminatively trained grammar, 93.71% of all tokens are labeled correctly (compared to 88.43% for a CMM) and 72.87% of records have all tokens labeled correctly (compared to 45.29% for the CMM).

Georgios Sigletos, et al (2005) have investigated the effectiveness of voting and stacked generalization -also known as stacking- in the context of information extraction (IE). A new stacking framework proposed that accommodates well-known approaches for IE. The key idea is to perform cross-validation on the base-level data set, which consists of text documents annotated with relevant information, in order to create a meta-level data set that consists of feature vectors. A classifier then trained using the new vectors. Therefore, base-level IE systems are combining with a common classifier at the meta-level. Various voting schemes are presenting for comparing against stacking in various IE domains. Well-known IE systems employed at the base level, together with a variety of classifiers at the meta-level. Results show that both voting and stacking work better when relying on probabilistic estimates by the base-level systems. Voting proved to be effective in most domains in the experiments. Stacking, on the other hand, proved to be consistently effective over all domains, doing comparably or better than voting and always better than the best base level systems. Particular emphasis also given to explaining the results obtained by voting and stacking at the meta-level, with respect to the varying degree of similarity in the output of the base-level systems.

Jordi Turmo, Alicia Ageno, and Neus Catal (2006) have described of online textual sources and the potential number of applications of knowledge acquisition from textual data has lead to an increase in Information Extraction (IE) research. Some examples of these applications are the generation of databases from documents, as well as the acquisition of knowledge useful for emerging technologies like question answering, information integration, and others related to text mining. However, one of the main drawbacks of the application of IE refers to its intrinsic domain dependence. For the sake of reducing the high cost of manually adapting IE applications to new domains, experiments with different Machine Learning (ML) techniques have carried out by the research community. This survey describes and compares the main approaches to IE and the different ML techniques used to achieve Adaptive IE technology.

Geoffrey E. Hinton, et al (2006) have presented how to use complementary priors" to eliminate the explaining away effects that make inference difficult in densely connected belief nets that have many hidden layers. Using complementary priors, he derive a fast, greedy algorithm that can learn deep, directed belief networks one layer at a time, provided the top two layers form an undirected associative memory. The fast, greedy algorithm used to initialize a slower learning procedure that fine-tunes the weights using a contrastive version of the wake-sleep algorithm. After fine-tuning, a network with three hidden layers forms a very good generative model of the joint distribution of handwritten digit images and their labels. This generative model gives better digit classification than the best discriminative learning algorithms the low dimensional manifolds on which the digits lie modeled by long ravines in the free energy landscape of the top-level associative memory, and it is easy to explore these ravines by using the directed connections to display what the associative memory has in mind.

Hal Daum and Daniel Marcu (2006) have proposed most basic assumption used in statistical learning theory is that training data and test data drawn from the same underlying distribution. Unfortunately, in many applications, the in-domain" test data is drawn from a distribution that is related, but not identical, to the "out-of-domain" distribution of the training data. He considers the common case in which labeled out-of-domain data is plentiful, but labeled in-domain data is scarce. He introduce a statistical formulation of this problem in terms of a simple mixture model and present an instantiation of this framework to maximum entropy classifiers and their linear chain counterparts.

Tak-Lam Wong and Wai Lam (2007) have developed a novel framework that aims at automatically adapting previously learned information extraction knowledge from a source Web site to a new unseen target site in the same domain. Two kinds of features related to the text fragments from the Web documents investigated. The first type of feature called as site-invariant feature. These features likely remain unchanged in Web pages from different sites in the same domain. The second type of feature called as site dependent feature. These features are different in the Web pages collected from different Web sites, while they are similar in the Web pages originating from the same site. In our framework, we derive the site-invariant features from previously learned extraction knowledge and the items.

Wenyuan Dai, Qiang Yang, et al (2007) have discussed traditional machine learning makes a basic assumption: the training and test data should be under the same distribution. However, in many cases, this identical distribution assumption does not hold. The assumption might be violate when a task from one new domain comes, while there only labeled data from a similar old domain. Labeling the new data can be costly and it would be a waste to throw away all the old data. In this paper, we present a novel transfer-learning framework called TrAdaBoost, which extends boosting-based learning algorithms (Freund & Schapire, 1997). TrAda Boost allows users to utilize a small amount of newly labeled data to leverage the old data to construct a high-quality classification model for the new data. He shows that this method can allow us to learn an accurate model using only a tiny amount of new data and a large amount of old data, even when the new data are not sufficient to train a model alone.

Chia-Hui Chang, et al (2010) have discussed the Internet presents a huge amount of useful information which is usually formatted for its users, which makes it difficult to extract relevant data from various sources. Therefore, the availability of robust, flexible Information Extraction (IE) systems that transform the Web pages into program-friendly structures such as a relational database will become a great necessity. Although many approaches for data extraction from Web pages have developed, there has been limited effort to compare such tools. Unfortunately, in only a few cases can the results generated by distinct tools directly compared since the addressed extraction tasks are different. This paper surveys the major Web data extraction approaches and compares them in three dimensions: the task domain, the automation degree, and the techniques used. The criteria of the first dimension explain why an IE system fails to handle some Web sites of particular structures.

3. PROPOSED METHODOLOGY

3.1 Proposed Approach

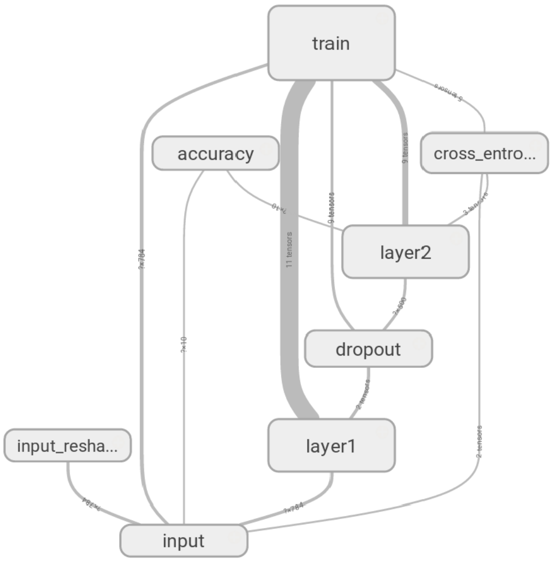

Figure 3.1 Proposed System Architecture

I have implemented the model on Python language The main purpose of this method is to propose an effective way to filter out valuable information from images and to reduce the complexity of the system. First we have analysed the system how accuracy varies with respect to iterations. For this purpose, In our system we have used cross entropy loss function and dropout regularization technique, in this technique we randomly dropout some weights, we have used dropout probability of .9 in training and testing for as expected there is as such no dropout. Figures in experimental and result phase shows the graph which represent how loss function varied with respect to iterations as expected at the end of iteration count. MNIST is a simple collection of Machine image dataset. The images in it are of handwritten characters identical. These images include labels for each and every image, for identifying what character it is. Here in above fig, the labels for the images are [5, 0, 4, and 1].

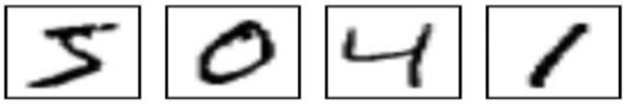

Figure 3.2: MNIST Image Representation

In this project, Model is being trained to look at images and guess what characters they are. Our purpose is to teach the model to achieve state-of-the-art performance. I have made the program to do it later! –here, we will begin with an exceptionally basic model, called a Multinomial logistic Regression. In this project, Model is being trained to look at images and guess what characters they are. Our purpose is to teach the model to achieve state-of-the-art performance -- I have made the program to do it later! –here, we will begin with an exceptionally basic model, called a Multinomial Logistics Regression. The MNIST information is part into three sections: 55,000 information purposes of preparing information (mnist.train), 10,000 purposes of test information (mnist.test), and 5,000 purposes of approval information (mnist.validation). This split is critical: it's fundamental in machine discovering that we have isolate information which we don't gain from with the goal that we can ensure that what we've realized really sums up!

As said before, each MNIST information point has two sections:a corresponding label and an handwritten character images and. We'll say the images "x" and the labels "y". Both the training set data and test set data contain images and their corresponding character labels; for instance the preparation pictures are train images if mnist that we have taken online and the preparation marks are trainlabels of train images character.

3.2 Deep Learning (DL)

Deep learning (deep machine learning, or deep structured learning, or hierarchical learning, or sometimes DL) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures or otherwise, composed of multiple non-linear transformations. Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc.. Some representations make it easier to learn tasks (e.g., face recognition or facial expression recognition from examples.

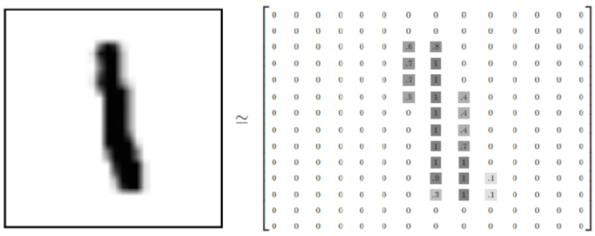

Figure 3.3: Image Array Representation

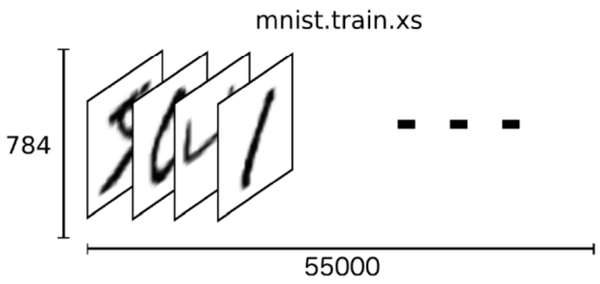

Figure 3.4: Vector Representation of image

3.3 Multinomial Logistic Regression (Multinomial logistic)

System need to be talented to look at an image and show the likelihoods for it being each character corresponding to the input image. For example, our System might look at a picture of a eight and be 85% sure it's a eight, but give a 10% chance to it being an nine (due to ambiguity of top loop) and it also matches with other character so we can never be 100% sure. This is a standard situation where a multinomial calculated relapse is a consistent, honest model. In the event that we need to dole out probabilities to a protest being one of a few unique things, multinomial strategic is the thing to do, in light of the fact that multinomial calculated gives us a rundown of qualities in the vicinity of 0 and 1 that mean 1. Indeed, even later on, when we prepare more advanced models, the last stride will be a layer of multinomial calculated. A multinomial strategic relapse has two stages: first we aggregate up the confirmation of our contribution for being in one of the many classes, and after that we change that proof into probabilities. To tally up the confirmation that a given picture is in a particular class, we do a weighted whole of the pixel powers. The weight is certain on the off chance that it is prove in support and negative if that pixel having a high power is confirm against the picture being in that class, and the accompanying outline demonstrates the weights one model scholarly for each of these classes. Red speaks to negative weights, while blue speaks to positive weights

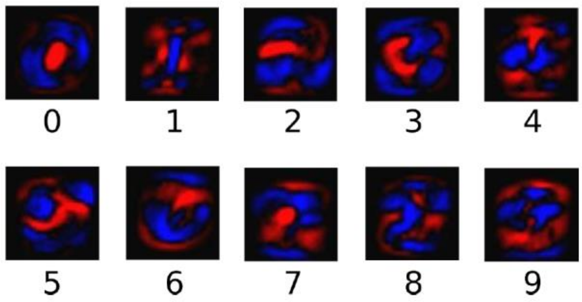

Figure 3.5. Multinomial Logistic Regression (Multinomial logistic)

Bias is some additional evidence added. Basically, we need to have the capacity to state that a few things are more probable free of the input information. Evidence for a class r for input x in the result is:

(—evidence‖= ΣjWr, jxj+br)……… ………(1)

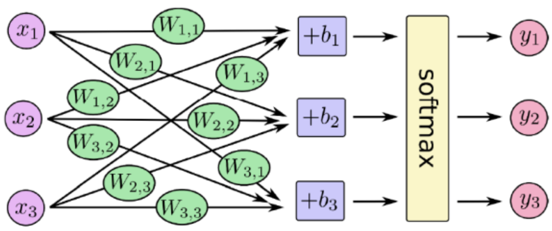

Figure 3.6: Softmax Regression

3.4 Training and Cross Entropy

With a specific end goal to prepare our model, we have to characterize what it implies for the model to be great. All things considered, really, in machine learning we regularly characterize what it implies for a model to be awful. We call this the cost, or the loss, and it speaks to how distant our model is from our coveted result We One extremely normal, exceptionally decent capacity to decide the departure of a model is called "cross-entropy." Cross-entropy emerges from contemplating data compacting codes in data hypothesis however it winds up being an imperative thought in loads of zones, from betting to machine learning. It's characterized as:

$$H_{y'}(y) = -\sum_i y'_i \log(y_i) \quad \text{………………………}(2)$$

Where y is our anticipated likelihood circulation, and y′ is the genuine conveyance (the one-hot vector with the character marks). In some unpleasant sense, the cross-entropy is measuring how wasteful our forecasts are for portraying reality. Comprehensively clarifying cross-entropy is past the degree of this wander, however it's well worth understanding.

3.5 Dropout

The task of prediction averaging is quite tedious for large neural networks. So to address this issue we use Dropout. In Dropout we usually off some of neurons from computation in hidden layer. With the end goal that these neurons can be kept from co-adjusting each other to much. On each feed-forward cycle, some of neurons present in hidden layer are separated from the system, Now they don't constitute as an individual from system. Separation of neurons happens just in hidden layer not in input and yield layers. So the sustain forward pass connected on changed system and after that data get back engendered on the same adjusted system. On every iteration weights and inclination are refreshed and dropout neurons are re-established. For next iteration we need to discover the following concealed layer of neurons to be detached on.

3.6 Deep Learning Algorithm

In machine learning in, a convolutional neural system (CNN) is a class of profound, bolster forward feed simulated neural system that have effectively been connected to dissecting visual symbolism. CNNs utilize a variety of multilayer perceptron's intended to require insignificant preprocessing.[1] They are otherwise called move invariant or space invariant manufactured neural systems (SIANN), in light of their mutual weights design and interpretation invariance characteristics.[2][3] Convolutional systems were propelled by natural processes[4] in which the network design between neurons is roused by the association of the creature visual cortex. Individual cortical neurons react to jolts just in a limited area of the visual field known as the responsive field. The open fields of various neurons halfway cover to such an extent that they cover the whole visual field. CNNs utilize moderately little pre-preparing contrasted with other picture characterization calculations. This implies the system takes in the channels that in conventional calculations were hand-designed. This autonomy from earlier information and human exertion in include configuration is a noteworthy favorable position.

3.7 Convolutional Neural Networks

It is distinguished from other neural networks by their superior performance with image, speech, or audio signal inputs. They have three main types of layers, which are:

1. Convolutional layer
2. Pooling layer
3. Fully-connected (FC) layer

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully-connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the image. Earlier layers focus on simple features, such as colors and edges. As the image data progresses through the layers of the CNN, it starts to recognize larger elements or shapes of the object until it finally identifies the intended object.

4. EXPERIMENTAL SET UP

Keeping in mind the end goal to prepare our model, we have to characterize what it implies for the model to be great. All things considered, really, in machine learning we normally characterize what it implies for a model to be awful. We call this the cost, or the misfortune, and it speaks to how distant our model is from our coveted result. We attempt to limit that blunder, and the littler the mistake edge, the better our model is.

4.1 Iteration Analysis and Accuracy Results

In our model we have found better accuracy as the number of iteration and less false positive results.

Figure. 4.1: Accuracy VS Iterations

This figure captures how accuracy varies with respect to iterations, red line represent accuracy numbers on test dataset and blue represents accuracy on train datasets. It can be seen that during training, there were fluctuations as gradient descent was trying to locate local minima for the initial iterations but as iterations count increased training accuracy curve got stabilized. 130th iteration was the point where saturation occurred from learning perspective after that there were improvements but quite small

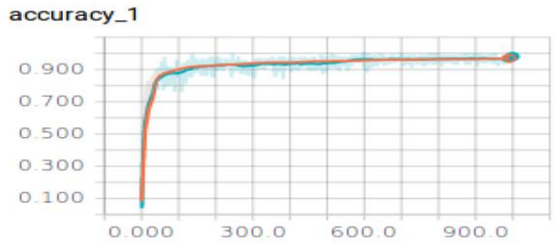

Figure 4.2 :Accuracy vs Iteration(2)

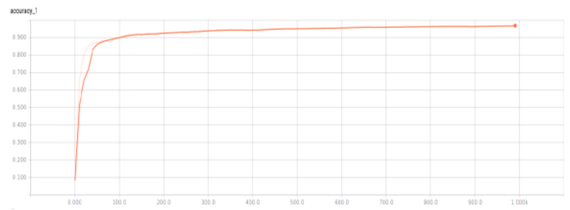

Figure. 4.3 : Test data accuracy VS Iteration

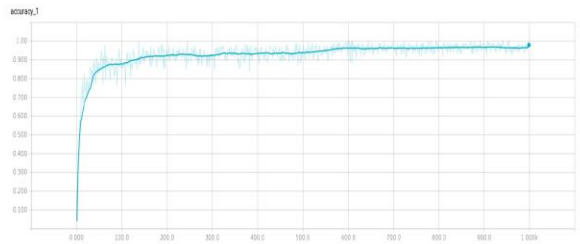

Figure. 4.4:Train data Accuracy vs Iteration

1.9 Cross Entropy And dropout Results

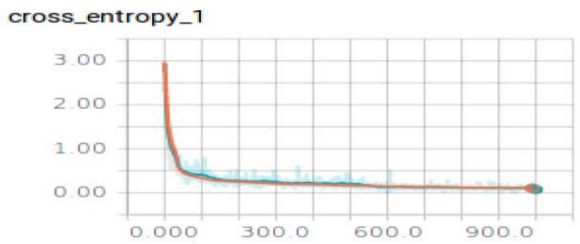

Figure. 4.5: Cross Entropy

We have used cross entropy loss function , this graph represent how loss function varied with respect to iterations as expected at the end of the iteration count error got minimized

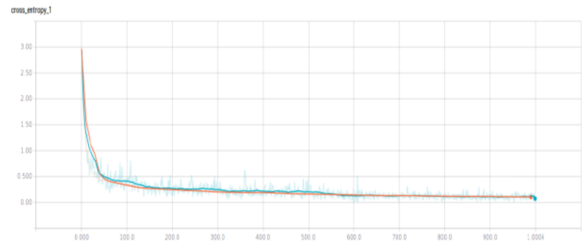

Figure 4.6 : Cross Entrophy(1)

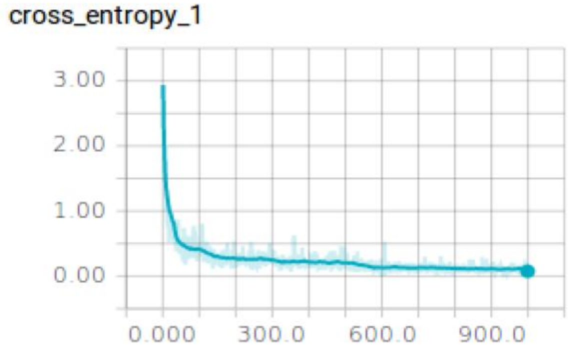

Figure 4.7 : Variation of Cross entropy loss on training data

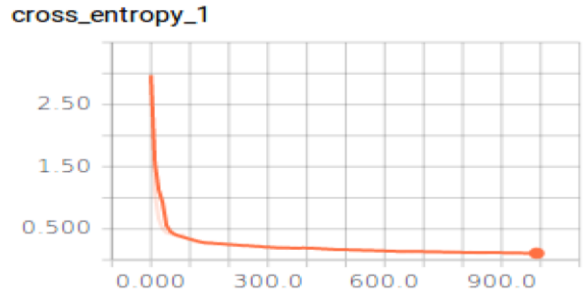

Figure 4.8 : variation of cross entropy loss on test data

Dropout deactivates each neuron probabilistically in both forward and backward pass. This method is substantially important in increasing generalization of the model.

Conclusion

In this we have seen that the model is adequately capable of correct detection in the domain of machine generated images and also training it with handwritten images, which we have not done for the purpose of our main work, would indeed yield a decent accuracy way more than 70%. The data gathered from MNIST for training our model contains all the varieties of images that we may encounter in real life scenario our model attained the accuracy of >92% which can increase with high system capabilities. Another viewpoint is this upgraded machine produced dataset owes its viability to the immense abilities of convolution layers. Neural systems are moving progressively and more toward End-to-End approaches and one of the primary ventures toward this way is highlight extraction stage, which normally restricts the extent of the model and was an unwieldy errand, is presently hindered utilizing convolution neural systems. In our research, we have successfully implemented neural network algorithms to classify large amounts of datasets to address the topic. There are various challenges for image analysis first. In an attempt to combat this, we have used neural network algorithm to classify the image made by slangs, misspellings and achieved accuracy of 93.145%. This paper tells about character recognition system for offline handwritten character recognition. The systems have the ability to yield excellent results. In this there is the detailed discussion about handwritten character recognize and include various concepts involved, and boost further advances in the area. The accurate recognition is directly depending on the nature of the material to be read and by its quality. Pre-processing techniques used in document images as an initial step in character recognition systems were presented. The feature extraction step of optical character recognition is the most important. It can be used with existing Character Recognition methods, especially for English text. This system offers an upper edge by having an advantage i.e. its scalability, i.e. although it is configured to read a predefined set of document formats, currently English documents, it can be configured to recognize new types. Future research aims at new applications such as online character recognition used in mobile devices, extraction of text from video images, extraction of information from security documents and processing of historical documents. Recognition is often followed by a post-processing stage.

REFERENCES

1. Nicholas Kushmerick “Wrapper induction: Efficiency and expressiveness” Artificial Intelligence (Elsevier), 118, Pages: 15–68, 2000.
2. Valter Crescenzi, Giansalvatore Mecca, and Paolo Merialdo “RoadRunner: Towards Automatic Data Extraction from Large Web Sites” Proceedings of the 27th VLDB Conference, Pages: 1-10, Roma, Italy, 2001.
3. Soumya Ray, and Mark Craven “Representing Sentence Structure in Hidden Markov Models for Information Extraction” Appears in Proceedings of the 17th International Joint Conference on Artificial Intelligence (IJCAI’01), Pages: 1-7, 2001.
4. William W. Cohen, and Wei Fan “Learning page-independent heuristics for extracting data from Web pages” TORONTO’99, Pages: 563-574, 1999.
5. Bing Liu, Robert Grossman, and Yanhong Zhai “Mining Data Records in Web Pages” ACM Int. Conf. on SIGKDD’03, August 24-27, 2003, Washington, DC, USA.
6. Bing Liu, Robert Grossman, and Yanhong Zhai “Latent Dirichlet Allocation”Journal of Machine Learning Research, 3, Pages: 993-1022, 2003.
7. Jason D. M. Rennie, Lawrence Shih, Jaime Teevan, and David R. Karger “Tackling the Poor Assumptions of Naive Bayes Text Classifiers” Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003), Pages: 1-8, Washington DC, 2003.
8. Oren Etzioni, Craig A. Knoblock, Rattapoom Tuchinda, and Alexander Yates “To Buy or Not to Buy: Mining Airfare Data to Minimize Ticket Purchase 170 Price” Int. Conf. on SIGKDD ’03, Pages: 2427-2435, August 2003, Washington, DC, USA.
9. Pranam Kolari and Anupam Joshi “webmining: research and practice, IEEE computing in science & engineering, pages: 43-53, july/august 2004.
10. Dragos Arotaritei, and Sushmita Mitra “Web mining: a survey in the fuzzy framework” Fuzzy Sets and Systems (Elsevier) 148, Pages: 5–19, 2004.

11. Valter Crescenzi and Giansalvatore Mecca "Automatic Information Extraction from Large Websites", Journal of the ACM, Vol. 51, No. 5, pp. 731–779.

September 2004.

12. Tak-Lam Wong and Wai Lam "A Probabilistic Approach for Adapting Information Extraction Wrappers and Discovering New Attributes" Proceedings of the Fourth IEEE International Conference on Data Mining (ICDM'04) Pages: 1-8, 2004.

13. Kyung-Joong Kim, Sung-Bae Cho "Fuzzy integration of structure adaptive SOMs for web content mining" Fuzzy Sets and Systems 148, Pages: 43–60, 2004.

14. Michael Azmy "Web Content Mining Research: A Survey" DRAFT Version 1, Pages: 1-15, November, 2005.

15. Paul Viola "Learning to Extract Information from Semi structured Text using aDiscriminative Context Free Grammar" Draft submitted to the conference ACM-SIGIR, Pages: 1-8, 2005.

16. Georgios Sigletos, Michalis Hatzopoulos, Georgios Paliouras and Constantine D. Spyropoulos "Combining Information Extraction Systems Using Voting and Stacked Generalization" Journal of Machine Learning Research, 6, Pages: 1751-1782, 2005.

17. Geoffrey E. Hinton, Simon Osindero, and Yee-Whye The "A fast learning algorithm for deep belief nets" Neural Computation, Pages: 1-16, 2006.

18. Qiang Yang, and Xindong WU "10 Challenging Problems in data mining research" International Journal of Information Technology & Decision Making, Vol. 5, No. 4, Pages: 597–604, 2006.

19. Ioan Pop "An approach of the Naive Bayes classifier for the document classification" General Mathematics Vol. 14, No. 4, Pages: 135–138, 2006.

20. Jordi Turmo, Alicia Ageno, and Neus Catal "Adaptive Information Extraction" ACM Computing Surveys, Vol. 38, No. 2, Article 4, Pages: 1-47, 2006.

21. Geoffrey E. Hinton, Simon Osindero and Yee-Whye The "A Fast Learning Algorithm for Deep Belief Nets" Neural Computation (Elsevier) 18, Pages: 1527–1554, 2006.

22. Hal Daum and Daniel Marcu "Domain Adaptation for Statistical Classifiers" Journal of Artificial Intelligence Research 26, Pages: 101-126, 2006.

23. A.Jebaraj Ratnakumar "An implementation of web personalization using web mining techniques" Journal of Theoretical and Applied Information Technology, 2005-2010 JATIT, Pages: 67-73, 2006.

24. Jun Zhu, Zaiqing Nie, Ji-Rong Wen, Bo Zhang, and Hsiao-Wuen Hon "Webpage Understanding: an Integrated Approach" ACM Int. Conf on KDD'07, Pages: 1-10, August 12–15, 2007, San Jose, California, USA.

25. Tak-lam Wong and Wai lam "Adapting Web Information Extraction Knowledge via Mining Site-Invariant and Site-Dependent Features" ACM Transactions on Internet Technology, Vol. 7, No. 1, Article 6, Pages:1-40, 2007.

26. Wenyuan Dai , Qiang Yang, Gui-Rong Xue, and Yong Yu "Boosting for Transfer Learning" Proceeding of 24th International Conference on Machine Learning, Corvallis, Pages: 1-8, 2007.

27. Insung Jung, and Gi-Nam Wang "Pattern Classification of Back-Propagation Algorithm Using Exclusive Connecting Network" World Academy of Science, Engineering and Technology, 36, Pages: 189-193, 2007.

28. Jen-Yuan Yeh, Hao-Ren Ke, and Wei-Pang Yang "iSpreadRank: Ranking sentences for extraction-based summarization using feature weight propagation in the sentence similarity network" Expert Systems with Applications (Elsevier)

35, Pages: 1451-1462, 2008.

29. Rajawat, A.S., Bedi, P., Goyal, S.B., Shaw, R.N., Ghosh, A. (2022). Reliability Analysis in Cyber-Physical System Using Deep Learning for Smart Cities Industrial IoT Network Node. In: Piuri, V., Shaw, R.N., Ghosh, A., Islam, R. (eds) AI and IoT for Smart City Applications. Studies in Computational Intelligence, vol 1002. Springer, Singapore. https://doi.org/10.1007/978-981-16-7498-3_10

29. Jason Weston, Frederic Ratle, and Ronan Collobert "Deep Learning via Semi-Supervised Embedding", Proceedings of the

25th International Conference on Machine Learning, Pages: 1-8,Helsinki, Finland, 2008.

30. Mohammed Salem Binwahlan, Naomie Salim, and Ladda Suanmali “Swarm Based Features Selection for Text Summarization” IJCSNS International Journal of Computer Science and Network Security, VOL.9 No.1, Pages: 175-179, January 2009.

31. Srivastava, S., Kumar, R.:(2013) Indirect method to measure software quality using CK-OO suite. International Conference on Intelligent Systems and Signal Processing (ISSP), Gujarat, pp. 47–51 (2013)

32. Rawat, R., Rajawat, A.S., Mahor, V., Shaw, R.N., Ghosh, A. (2021). Surveillance Robot in Cyber Intelligence for Vulnerability Detection. In: Bianchini, M., Simic, M., Ghosh, A., Shaw, R.N. (eds) Machine Learning for Robotics Applications. Studies in Computational Intelligence, vol 960. Springer, Singapore. https://doi.org/10.1007/978-981-16-0598-7_9.