ISSN: 0474-9030

Our Heritage

Vol-68-Issue-15

January-2020

UGC Care Listed Journal

Survey Of Receipt Identification And Classification Using Machine Learning

[1]Garima Ghatge, [2]Vedant Ghavate, [3]Prithvi Chaudhari, [4]Prof. P. B. Wakhare
Information Technology
AISSMS Institute of InformationTechonlogy
Pune,India

***Abstract*:**

The automation of identifying and classifying receipts requires a sophisticated which utilises various technology and algorithms. Firstly, images are cleaned with image processing technique like rasterization, binary black and white classification and skewing. The tesseract engine is used in the process of Optical Character Recognition, to convert image into virtual text. Tesseract is a powerful engine which uses multiple algorithms to enhance accuracy. Thirdly, the generated text is used to extract meaning by defining extraction rules and building a classifier based on predefined entities. By using these methods meaning and semantics of the receipt are extracted. This data is extracted and the meaning of data is to be stored in a non-relational database management system as the receipts are not homogeneous in nature. Multiple invoices received will be stored in separate collections for building a huge record. The user is expected to upload a month expenditure on this database. Lastly, a report based on aggregation and generalization of this data is to be created. The report has to be specialized to the user's requirement and input using various data visualisation tools in this way we aim to automate the process of managing and classifying receipts.

Keywords: *Receipt, Optical Character Recognition, classification, invoice*

I. INTRODUCTION

There is a growing need for maintaining a tab on all monthly expenditure by using bills or receipts. Since most of the bills or receipts are paper based, it is impractical to maintain the bills as the ink on the bills fades after a few months. The user does not get an overview of the aggregated data and has no easy way to know his/her expenditure at a glance. Our project aims to automate the entire process using machine learning techniques, with mobile hardware improving day by day and the computational power of Cloud Platforms we can build a system to facilitate this goal.There are many techniques for Optical Character Recognition, but we need to use specialised algorithms that are designed specifically to provide accurate results on invoices. It is also necessary for this specialised OCR to be available to all users just by clicking a picture of the bill. The text extracted from the bill including various numeric details like shop name, amount, date that are used in building a database. After understanding the contents of the bills, we need to deliver an expense report. Generating an expense report is another important aspect of our project, the data collected from all the bills is to be classified and categorised. A neural network will categorise items based on their utility, furthermore these categories along with the database will help build coherent report for the user.

Existing Work Carried in Document Identification and Classification Techniques:

A. Building text features for object image classification

The methodology used in “Building text features for image classification” [8] primarily focuses on a text based image feature and shows that it significantly improves performance on hard object classification problem. An auxiliary dataset is downloaded from the internet which has images that are annotate with tags. The tags are not checked or checked and it is expected that they will be noisy. Millions of images that are available on the internet are not labelled but the text that is associated with the images provides a much better understanding towards image analysis. It is built on two insights, firstly it is often simpler to determine the content of the image based on the text surrounding it. The second insight when matching with the simplest image feature very similar images are found given a huge dataset. The strategy that is most commonly used to improve annotation quality is to gather a new collection of images that will be used as training data.The features used in this method for object image classification are Visual features, Internet dataset, Text features, a Classifier and Fusion of the first two classifiers present. The text features are relied upon highlighting semantic importance of the images and provided a straightforward way to image analysis. A dataset containing test and training images is present. A dataset of images downloaded from the internet is present which has surrounding text associated with it.

Visual features are used to train the visual classifiers and find the images that are the nearest neighbors. Five types of visual features are used namely SIFT, Gist, Color, Gradient and Unified. The SIFT feature is used for recognition of objects and image matching. Scene categorization is done by using the Gist feature. Each image is described in 960 dimensions. Color features are extracted in RGB space. In Gradient feature the image is represented as 256 dimension vector. A concatenation of the four features above is Unified feature. The association that is present between the images and text is provided by the internet dataset. The images are represented in a way which more directly highlights the semantics is done in Text feature. Various classifiers can be used, SVM classifier is used in this approach. One feature can be unreliable so fusion of two classifiers is done to form a third classifier.

Visual features are extracted from each training image to find its K nearest neighbor image from the dataset downloaded from the internet. The text that is associated with the internet images that are near neighbor is used to construct the text features. The same procedure is followed for the test image to build it’s text features and use it to predict category labels. A separate classifier is trained on the visual features. The final prediction is formed by a third classifier which gets values from the visual and text classifiers.

The text that is produced by matching a large dataset of images having noisy annotations to an image provides a powerful feature that gives positive results.

B. Automated Receipt Image, Identification, Cropping and Parsing

The methodology used in “Automated Receipt Image, Identification, Cropping and Parsing” [4] mainly focuses on extraction of date and price from the receipts. The receipts need not be of a single format and vary widely from simple grocery bills to complex air line ticketing receipts. The concept of receipt localisation and cropping is applied to segregate the background image from the actual actual image of a receipt. Every receipt is rectangular in shape and the process of detecting the edges is considered to be the most crucial while separating it from a complex background. The image is transformed from RGB color scheme to a grayscale for speeding up the process. This approach makes use of three edge detection methods to improvise its accuracy; Line Segment Detection, Probablistic Hough Transform and Holistically-Nested Edge Detection. Line Segment Detection algorithm is based on Guassian pyramids to rectify line support regions. The line support regions are further filtered to extract the sharp edges in the image irrelevant of its length and width. Probablistic Hough Transform is another edge detection technique. A 5x5 Guassian kernel is applied to a gray scale image to filter out the noise. The processed image is used for edge detection by Canny Edge Detector. The Probablistic Hough Transform then extracts the leading edges from an array of edges by setting a minimum threshold length.A deep learning model is applied in Holistically-Nested Edge Detection to rectify the most significant edges in an image. However, this method is not advantageous when the background image is noisy.

In order to detect the most likely corners of the receipt the edges detected by the three detection techniques undergo affine transformation and unwrapping. The corners are grouped by using various combinations. Each combination in a group is assigned a score, which is dependent on the area of the receipt covered by these corners and the sharpness edges between the corners must be nearly 90 degrees. The group with least score is chosen as the most likely corners of the receipt. The edges between these corners is axis-aligned and are used for cropping out the receipt from its background for further processing.The system makes use of Tesseract, an engine provided by google to detect the characters on the receipt. The processed image is fed as input to Tesseract to extract the textual data from the receipts. The system returns the date and amounts on a receipt as output. The system was able to classify 36/50 receipts correctly.

However, the work suggested in this paper does not focuses only on the extraction of the date of purchase and the prices, it does not suggest significant methods for extraction of other details in the receipts. Also, it does not focus on the classification of receipts after extracting the data from the receipts.

C. A Robust Algorithm for Text Region Detection in Natural Scene Images

In this paper [1], a system is proposed for detection of text region in the scene images. To extract the text from an image it is first important to identity the text region first which is done using support vector machine(SVM). Then the text is extracted from the text regions based on certain size and alignment constraints. In this paper, a novel method is used for extraction of text regions in scene images by means of new colour feature and support vector machine with wavelet moments. This method is intended towards the robustness to the variations in the alignment, illuminations, text, scale, etc. The new colour feature used in this method is mainly to contribute to the robustness to changes in illumination, and the wavelet moments contribute to the variations of scale. The red, green, blue (RGB) components are analyzed first and then the colour feature is selected. After this the clustering algorithms are applied. The clustering segments the scene image into objects, which are then applied to a support vector machine. The support vector machine indentifies whether the object is a text or not. If it is indentified as a non-text region it is discarded. This method detects the text individually as they belong to the segmented object. In this method, there is detection of text directly from images without analyzing the entire image. Hence, it is efficient and robust to variations in illumination and changes in scale of the text.

The proposed method has the following steps: 1) Image segmentation - pixels are divided inti achromatic and chromatic regions after analysis of RGB. Then clustering algorithms are applied to segment the objects in the image. 2) Feature extraction - By the wavelet transform the frequency characteristics of an object are extracted. 3) Training & text object detection - support vector machine (SVM) is trained with features to detect the text objects in the image. The method proposed in this paper is a combination of colour representations. It includes quantization of colour with edge detection, RGB and HSI combined representations, RGB representation based colour quantization. The k-means clustering algorithm is applied to every pixel for text segmentation. In this paper, the computation time is drastically reduced by using histogram for k-means algorithm. For hue values [0.6 - 1.0], k-means clustering algorithm considering cyclic property is used. For intensity values [0.0 - 0.4], k-means clustering algorithm considering euclidean distance is used. In wavelet based feature extraction algorithm the input is a segmented image with a gray scale tone. The image is then size-normalized into 64x64 image. Divide it into 4x4 regions and from these blocks compute the first and second moments which are then taken as features.

In text area classification based on SVM, two classes are separated with a decision surface which has the maximum margin. In this paper, only a binary classification task is considered with n labelled training instances. There is a hyperplane that separates the training instances to classify areas as text or non-text regions. Therefore, in this system the approach is split into three parts: natural images segmented using clustering, feature extraction, text region detection using SVM. The effectiveness of the proposed method is proved by the experimental results that locate most of the text regions in the test images. This approach does not give accurate results for images that are complex in terms of uneven lighting and background.

D. Text Detection and Recognition in Natural Scene Images

In this paper [2], a system is proposed for detection, extraction and recognition of text in scene images. Text extraction from image is divided in 4 stages: 1) text detection, 2)text localization, 3)text extraction, 4)text recognition. The system is proposed for detection of text and localization of text in natural scene images as it is challenging because of fluctuation in size, colour, orientation, alignment and is affected by image distortion, background, degrading, etc. In the initial stage, image is prepressed to detect the text region in the image using a feature descriptor that is histogram-oriented gradient. Then segmentation that is local binarization is applied.

In the text extraction stage, parameters like normalized width height ratio, compactness are considered to identify the text and non-text components. In the last stage, text is recognized using the zone centroid and the image centroid-based distance metric feature extraction. In the text region detection stage, the image is converted to a gray level image. The aim of the detection of text region is not to predict exact text position but to predict the probabilities of text position and scale information. Histogram of gradients divide the image in various connected regions called as cells and compute the histogram of gradient edge orientation for pixels of the cell.

In text extraction stage, the connected text components are extracted from the text line in the image and this can be used to localize the text in the image accurately. In this paper, normalized height and width ratio along with compactness are the two parameters used to identify the text and the non-text components accurately. In pre-processing step, the normalization and thinning are important steps. The normalization of the image converts the characters into 50x50 size as the size varies from font to font and thinning reduces the thickness of each text line pattern to a single pixel. The zone based feature extraction is used for the extraction of the characters.

Zone based feature extraction gives appropriate results even when certain pre-processing steps are not performed such as smoothing, filtering, etc. Algorithm one provides feature extraction system based on image centroid zone (ICZ). The image is equally divided into 50 parts. The image centroid is calculated and the average distance from the centroid to every pixel of the image is calculated. As a result, n features will be provided for recognition. Algorithm two provides feature extraction system based on zone centroid zone (ZCZ). In this, the zone centroid is calculated and the average distance from the centroid to every pixel is calculated. This is done for all zones. As a result, n features will be provided for recognition.

For evaluation purpose a name plate image is taken as the input image and HOG is used for text region detection and segmentation.The disadvantage of using HOG is that it is not scale and rotation invariant.Text is extracted and recognized using ICZ and ZCZ algorithms. 90% accuracy is achieved for text extraction and recognition.

E. A survey of Feature extraction and classification techniques in OCR

The process of identifying characters from an image has been an area of research for the past few years. The work depicted in “A survey on Feature extraction and classification techniques in OCR” [3] is one such example. Optical Character Recognition is a science of extracting textual data of interest from a digital image. The study in this paper is based upon three main aspects segmentation, feature extraction and classification. Every image is considered to be composed of dots also known as pixels. These pixels are arranged in a particular form to depict shapes and in turn represent complex pictures. The pixels are assigned different RGB values. Every image is thought to be of m pixels in the vertical direction and n pixels in the horizontal direction. In this OCR methodology the image undergoes image processing, segmentation, feature extraction and classification.

In the image processing stage, the digital image is scaled to an appropriate value to ensure readability of the document. If an image is captured in low light, it becomes difficult for the Optical Character Recognition Engine to understand the text, hence to enhance the image quality increase the contrast between the image foreground and background. Many OCR make use of a monochromatic color scheme, that is it converts a multi-colored image to a white and black image. Many a times, people do not align the image at proper angles, which makes it difficult for an Optical Character Recognition Engine to extract text from the document. In this case, the image is de-skewed to align all textual data horizontally from left to right.

The processed image undergoes segmentation. Segmentation is a process of grouping pixels in an image, such that each group of pixels denotes some specific part of the image. Each pixel in the digital image is assigned a label. The grouping of pixels is done based on these pixels for the purpose of showing similar characteristics. The grouping of pixels is known as segments. The segments of an image are classified into lines, words and then characters. Pixels in a segment show similarity in terms of color ,intensity, texture or a combination of them.

Every character has unique features with respect to the other characters. The process of identifying these unique features is known as feature extraction. The features are broadly classified as global or statistical features and structural or topological features. The global features focus on the arrangement of the pixels in a character matrix. They can be easily detected in comparison to topological features. They use methods like calculating the distance of the pixels from the centroid of a character matrix(moments), dividing the matrix into zones(zoning) and calculating the number of black and white pixels(projection histogram). The structural feature extraction focuses on the geometry of the characters. It takes into consideration the curves and edges of a character, the number of a strokes in a character, crossing of strokes and the end points of a character.

Classification is the most crucial phase in Optical Character Recognition. It selects a subset of features from the features extracted in the previous phase. The syntactical method classifies characters into classes based on its components and the relationship amongst them. In template matching, classification is done by matching the matrix of a character with a predefined character which is accurate. An Artificial Neural Network can also be applied to extract the important features of a character.

The Optical Recognition Engines go through these phases to ensure high accuracy in prediction of characters in an image. Optical Character Recognition can be applied to various forms of documents.The study does not suggest any platforms or engines for Optical Character Recognition which can be directly incorporated.

F. Automatic Text Categorization Using Neural Networks

This paper [5] demonstrates the results from the experiments in a semantic categorization of MEDLINE articles. The main goal of this research is to build a neural network and to train them in assigning MeSH phrases based on term frequency of single words from title and abstract. The experiments compare the performance of a counter propagation neural network to a backpropagation neural network. Results obtained by analysing a set of 2,344 MEDLINE documents are used to conclude the use of the neural network to category documents based on vocabulary

The data set used for training and testing consists of 2,344 Medline documents described in Hersh, Hickman and Leone (1992) and used by several researchers. Each document of this collection includes the title, authors, citation information, abstract and a set of manually assigned MeSH phrase.

This collection was processed using the SMART system. The processing consisted in tokenizing words from title and abstract, eliminating common words using a stop list, stemming remaining words, and computing the frequency of each stem in each document. MeSH phrase assigned by indexers from the NLM consist of phrases that can include qualifiers (i.e., cancer, DIAGNOSTIC). For this study, the qualifiers were separated and considered as independent MeSH phrase.

The next processing step consisted in identifying the unique stemmed-words from titles and abstracts and the unique MeSH phrases from the entire collection and computing each element's document frequency. A total of 12,292 stemmed-words and 4,049 different MeSH phrases were found in the entire collection. The document frequency varies from 1 to 1,511 for word-stems and from 1 to 2,102 for MeSH phrases.

The size of the input of the neural network depends upon the number of stemmed words. Thus, it was decided to reduce this size by setting a minimum document frequency threshold on the entire collection. Phrase with higher document frequency usually correspond to general phrase, while phrase with very low document frequency correspond to very specific or rare phrase.

They selected a threshold of 30 for document frequency since it offers a reasonable reduction in the number of phrase while retaining phrases that are not too general. Using this criterion, the number of possible nodes in the input layer was thus reduced to 1,016 stemmed-words.

Once the size of the input and output layers was decided, the size of the intermediate layer was chosen as 3 times the output of the Grossberg layer. The reason is that it is assumed that this layer represents a level of abstraction from the words in the title and abstract of documents to the more general MeSH concepts. The proposed neural network then has three layers of 1,016, 540 and 180 nodes. The middle layer is the output of the Korhonen layer and the input for the Grossberg layer.

The collection of 2,344 documents was divided in two : 586 training documents md 1,758 documents for testing. The training documents were chosen randomly.

The counter propagation network was trained using the 586 documents. For training the Kohonen layer, the criterion was set of that the average distance between patterns was less than a predefined value or that the number of cycles was less than a predefined maximum. An initial run was trained for a maximum of 25 cycles. The average distance per pattern in the last iteration of the Kohonen layer was 0.75 and the error per pattern in the last cycle of the Grossberg layer was 19.15. They also trained a backpropagation network using the same data and three layers of the same size as the counter propagation network. The backpropagation network converged to the value of acceptable error (0.2) in 19 cycles and took about 6 hours.

The results obtained in this work suggest that neural networks could be an important tool in automatic text categorization. Given proper input, the network learns to assign categories of a controlled vocabulary using free text from the title and abstract. The backpropagation network performed better than the implementation of Counter propagation networks. Counter propagation network, not considering its speed, is a more feasible solution because the knowledge obtained by the network during the training phase can be used to frame fuzzy rules, which can be used to establish the control vocabulary. This method still cannot be applied on a diverse dataset like a grocery database as the control vocabulary would to vast to establish manually.

G. A Table Extraction from Document Images using Fixed Point Model

The paper [6] presents a fresh approach to learning-based framework to identify tabular data from scanned document images. The approach is designed as a structured labelling problem, which learns the layout of the document and labels its various entities as table header, table trailer, table cell and non-table region. It illustrates development features which encode the foreground block characteristics and the contextual information. These features are provided to a fixed-point model which learns the relationship between the blocks. The fixed-point model obtains a contraction mapping and assigns unique label to each block. The fixed-point model captures the context information in terms of the neighbourhood layout effectively. To automatically extract this information by digitization of paper documents, the tabular structures need to be identified and the layout and inter-relationship between the table elements need to be preserved for subsequent analysis.

This paper shows significance the layout of a document image by extracting the attributes of foreground and background regions and modelling the correlations between them. Using these attributes, a fixed-point model captures the context and learns the inter-relationships between different foreground and background document entities to provide them with an unique label which can be, table header, table trailer, table cell and non-table region. Regions which get table related labels are clustered together to extract a table.

The Fixed-point model captures the neighbouring information and models the correlation between the different nodes to predict the label of each node. The fixed-point model utilizes the context information and obtains a contraction mapping to assign a unique label to each element of document image.The proposed method for table detection is tested on a dataset of 50 images which were picked from UW-III dataset, UNLV dataset and a newly collected dataset consisting of documents with different table layout.An overall accuracy of around 96% for assigning a label to each document region. Table-cells and non-table blocks are labelled with high accuracy due to their significant appearance and contextual features. This approach gives correct labelling for tables present in different page layouts. Thus, it is clear that the method learns the layout with effectively and can be used on images with complex layouts.This paper presents a revived approach to learning-based framework for the problem of layout analysis and table detection in document images with complex layouts. An amalgamation of foreground and background features is extracted and used with the fixed model for learning the context information. This helps in learning the document layout and labelling the different regions. The experimental results are promising and this approach works well on a heterogeneous collection of documents. The devised method is general and does not rely on heuristic rules such as the presence of horizontal and vertical lines. It provides an alternate to most of the present rule-based and learning-based systems. However, the system is based on visual cues of lines and dotted separation of labels, this severely limits the application of the system of multiple documents.

H. Automatic image tagging via category label and web data

In this paper "Automatic Image Tagging Via Category Label and Web Data" [7] the methodology used primarily focuses on assigning each image a category label so it can automatically recommend other related tags to the image which in turn reduces the human annotation efforts. The method used in this paper requires users to annotate each image with a single category label first. Classifiers are constructed using these category labels to get rid of the irrelevant web images. A small subset of the relevant images is selected from the codebook using sparse coding technique. On the other hand the corresponding tags of the relevant images form a tag pool. The local and global agglomeration technique is used to automatically select and recommend the most suitable tags to the image which is to be tagged. The human intervention is greatly is minimized in the image tagging process.

This method makes use of three modules Noisy Images Filtering to filter out the irrelevant web images, Relevant Image Selection to select images which are significantly related for tag propagation and Tag Re-Ranking to re-rank most of the tags which are corresponding to relevant images and make use of the most appropriate tags for image tagging.

Noisy Images Filtering makes use of Spatial Pyramid Matching (SPM) to reduce the level of noise in the images and also reduces the number of web images by re-ranking these web images according to the decision values given by the SVM classifier. It makes use of the Histogram intersection kernel and the images that have higher decision values are more suitable images for the given category. Relevant Image Selection makes use of Sparse Coding for image annotation, selecting web images that are relevant and label propagation. Sparse Coding uses different parts from different images to construct the objective image linearly. The noise level is reduced to some extent in sparse coding as the ranking of relevant image set is not considered. Tag Re-Ranking uses WordNet to correct the words that are misspelt, the irrelevant tags are then filtered out and also the tags which have frequency less than 2 are removed. The global ranking based assignment tag is used which increases the accuracy than local ranking bases tag assignment.

Experiments carried out to evaluated the method had 8 categories namely badminton, bocce, croquet, rock-climbing, polo, rowing, sailing and snowboarding. 137-250 images were selected for each category and 1792 images in all. The images whose resolution was higher than 2 or less than 0.5 were eliminated. The first 1500 images of the output of SVM classifier were used as the codebook images. The top 3 recommended tags are all correct for most of the categories. The performance for some of the categories is relatively low due to noisy images. Each image is assigned a category before and the related tags are automatically tagged to this image by making use of sparse coding. The human intervention is reduced greatly in this way. It is seen that the sparse coding based methods provide much better performance than knn based method and which in turn proves the effectiveness of sparse coding.

II. CONCLUSION

This paper presents the existing work done on identifying features and classification on semantic relationship of words in a document and images with their captions. The paper is focused on techniques used in classification of images and textual data and extraction from tabular data. However, it is observed that there is no concrete system designed to extract data from a receipt-based document. Some research shows that an amalgamation of the techniques can be used to make classification model for the data extracted from the receipt. Thus, it motivates the purpose of designing a robust system for serving the purpose and overcoming the limitations and inadequacy of the existing work.

REFERENCES

[1] Jonghyun Park and Gueesang Lee,"A Robust Algorithm for Text region Detection In Natural scene Images.", can.j.elect.comput.eng., vol.33, no. ¾, summer/fall 2008

[2] Pise, A., &Ruikar, S. D. (2014, April),"Text Detection and Recognition in Natural Scene Images." In Communications and Signal Processing (ICCSP), 2014 International Conference on (pp. 1068-1072). IEEE.

[3] Rohit Verma and Dr.Jahid Ali, "A-Survey of Feature Extraction and Classification techniques in OCR Systems." Proceeding of the international journal of Computer Applicationand Information Technology, Volume 1, Issue 3, November 2012.

[4] Alex Yue," Automated Receipt Image, Identification, Cropping and Parsing"

[5] Miguel E. Ruiz, Padmini Srinivasan, "Automatic Text Categorization Using Neural Networks"School of Library and Information Science, The University of Iowa

[6] Anukriti Bansal, Gaurav Harit, Sumantra Dutta Roy, "Table Extraction from DocumentImages using Fixed Point Model"

[7] Wang, G., Hoiem, D., Forsyth, D. (2009). Building text features for object image classification. 2009 IEEE Conference on Computer Vision and Pattern Recognition.

[8] Gao, S., Wang, Z., Chia, L.-T., Tsang, I. W.-H. (2010). Automatic image tagging viacategory label and web data Proceedings of the International Conference on Multimedia -MM '10