Content based Caption Generation for Images Embedded in News Articles

Amitkumar Kohakade
Post Graduate Student at Pune Institute of Computer Technology, Pune

Emmanuel M.
Head of Department of Information Technology, Pune Institute of Computer Technology, Pune

ABSTRACT

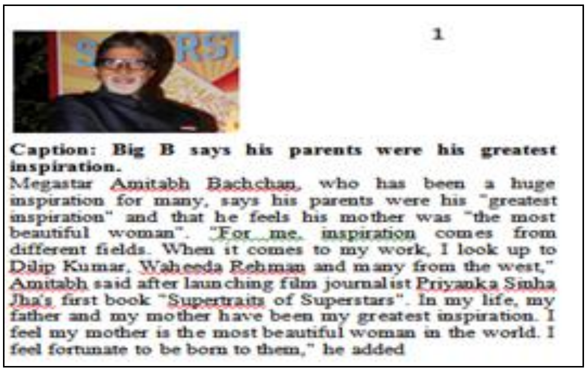
In current digital world Content based Image retrieval is becoming critical problem as size of data on Internet increasing rapidly. When the image is embedded in news article it is retrieved by manipulating words annotated to that image, text placed surrounding to that image etc. Many times this annotation, caption generation is done manually. It reduces accuracy, increases time span and makes it as tough task. We proposed a new approach for generating caption for such images. Approach presented here focuses on important terms occurring in news like named entities, using term weighting find out weighted terms which helps in describing news. On other hand by image processing we find out who's in picture as it helps in making accurate caption by using face recognition and it will increase image retrieval. Some of experiments presented here shows performance of face recognition algorithms on standard datasets and also on own developed face dataset, also we train NER model on Indian names which gives better results. As it covers text and image content it helps in generating better caption and also for improving image retrieval accuracy.

Keywords

Caption generation, Name entity recognition, Text Processing, Face Recognition.

1. INTRODUCTION

As days go on digital data on Internet like image, audio, video increasing rapidly. Searching data in this huge database is now becoming big problem for users. Search engines play an important role in finding results for user query like retrieving image or site search etc.. When user search for particular image search engine checks different sites, images words annotated to that images, also search words places surrounding that image. Images which match to user query by these methods mostly retrieved by search engine. As here it just check something which is sometimes irrelevant and retrieves image on that basis without actually knowing image content, many times this approach reduces the accuracy of image retrieval. Annotation, caption like methods which decides retrieval accuracy must be correct and it describe that image overall by considering image along with other content into consideration. Our objective is to generate a caption for images which are embedded in news articles by processing text placed near that image and actual image content by finding who's in picture? As many times this is done manually which increases labour task and time for generating exact caption. Usually images are annotated with keyword which helps to simplify access to them. Many times this search results contains irrelevant images. A lot of research is going on generating descriptions for image automatically. There are different machine learning approaches which learns first, from given training samples and then used for image annotation, caption generation etc. [26] Presents model which learns semantic of images which helps in automatically annotating image with keywords. Research is going on generating automatic description for image which helps in content based image retrieval. This also helps journalists for content writing or in case of finding images related to their article. With this it also increases accessibility of multimedia data on Internet for visually impaired people [17]. In general basically in two steps caption is generated. Initially find out what the news and image is saying and then in second step say same thing in short. Fig. 1 shows template of news, caption and image taken from NDTV bollywood news site.



Caption: Big B says his parents were his greatest inspiration.

Megastar Amitabh Bachchan, who has been a huge inspiration for many, says his parents were his "greatest inspiration" and that he feels his mother was "the most beautiful woman". "For me, inspiration comes from different fields. When it comes to my work, I look up to Dilip Kumar, Waheeda Rehman and many from the west," Amitabh said after launching film journalist Priyanka Simha Jha's first book "Supertraits of Superstars". In my life, my father and my mother have been my greatest inspiration. I feel my mother is the most beautiful woman in the world. I feel fortunate to be born to them," he added

Figure 1 Image, Caption and News taken from NDTV Bollywood news domain

2. LITERATURE SURVEY

A Caption helps in retrieval of images which are embedded in the news document. As it decides the accuracy of image retrieval, it must consider image and document content. Jia-Yu Pan et. al. introduced caption generation method based on correlation between image features and keywords [15]. Recently Feng et. al. presents machine learning approach for caption generation, in which model learns from database of news articles, images in that article and captions for them while training [27]. Image annotation is the task in which image is tagged with keyword or captions which help in image retrieval. [14] Assumes regions in an image can be described using vocabulary of blobs and these blobs are generated from image features using clustering. This blobs works as an annotation for image. Allan Hanbury et. al. discusses three image annotation approaches free text

Table 1 Different term weighting methods

Paper Name	Published Year	Method	Algorithm	Description
Supervised and Traditional Term Weighting Methods for Automatic Text Categorization [19]	2009	TF	W(d,t) = TF(d,t)	the number of times term occurs in document
		IDF	IDF(t) = log $\left(\frac{N}{n}\right)$	n be number of document in which term t occurs N be total number of document in collection
		TF-IDF:	W(d,t) = TF(d,t).IDF(t)	
		CHI square :	X^2(f,c) = $\frac{N*(AD-BC)^2}{(A+C)*(B+D)*(A+B)*(C+D)}$	A be the times both feature f and class c exists, B be the times feature f exists, but class c doesn't exist, C be the times feature f doesn't exist, but class c exists, D be the times both feature f and class c doesn't exist, N be the total number of the training samples.
		Term Frequency CHI square	TF . X^2	
		Term Frequency-Relevance Frequency	TF * log $\left(2+\frac{a}{\max(1,c)}\right)$	a is number of document contain positive category term c is number of document contain negative category term When there is no document which contains negative category term then it is consider as 1 for avoiding divide by zero case
Graph-based term weighting for information retrieval [23]	2012	Graph Ranking Weight: Text Rank	$S(v_i)$ = (1-Ø) + Ø $\sum_{j\epsilon V(v_i)}\frac{S(v_j)}{\lvert V(v_j)\rvert}$ (0<= Ø <=1)	S(v_i) and S(v_j) denote the score of vertex v_i and v_j $V(v_i)$ and $V(v_j)$denote the set of vertices connecting with v_i and v_j Ø is a damping factor
		Link Based weight: Text Link	$S(v_i) = \delta(v_i)$	$\delta(v_i)$ is the average degree of a vertex
		Graph ranking weight: PosRank	$S(v_i)$ = (1-Ø) + Ø $\sum_{j\epsilon In(v_i)}\frac{S(v_j)}{\lvert Out(v_j)\rvert}$ (0<= Ø <=1)	$In(v_i)$ - all vertices linking to v_i $Out(v_j)$ - all vertices linking to v_j
		Link based weight: PosLink	$S(v_i)$ =| In (v_i)|	In (v_i) - vertex in-degree

annotation, keyword annotation and annotation based on ontology [2]. Neural networks have powerful information processing ability and simple implementation so it has been applied to image annotation problems. With the help of it Zenghai et. al. proposed an adaptive recognition model for image annotation which consists of classification network and correlation network [28]. These neural networks reduce time for annotation and provide advantages over concept association network. Many times these annotated words rarely match with user query and it degrades retrieval accuracy. In this paper we consider news document, find out important terms by term weighting. In domain of text summarization a lot of research has done yet and also going on. Above table 1 give overview of some term weighting methods. Many times text document is represented in form of vector space model. There are number of term weighting methods like TF, TFIDF, TFRF etc. are widely used in text processing. Also sometimes graph based term weighting methods are used for sentence ranking which are also helps in finding important terms, sentences. Saurabh et. al. introduced new method for term weighting as Positive impact factor. Positive impact factor of feature is used to calculate its negative impact for other category that helps in text categorization [21]. Michel et. al. presents statistical models for term selection, term ordering for headline generation task [20].

In text summarization named entities helps a lot as these are nothing but like person name, location, organization name, date, time, money etc. Here as we are going to generate caption for image which sometimes contain people faces, name of those people must come into caption, so name entity finder helps in that case. There are many name finder models available like OpenNLP's name finder, Stanford's NER, GATE's NER, many of them are works better for European names than Indian names. So for it we train OpenNLP's NER model for Indian names.

On the other hand image processing plays an important role as caption must describe image content. In computer vision lot of research has done and still going on. [27] Represents image as bag of visual terms. Using SIFT algorithm SIFT descriptors are find out and then K-means algorithm is applied for finding descriptor set of visual terms. Image may contain different object like car, tree, human faces, caption must say about it. A lot of research is going on for automatic description generation for images by segmenting that image, then finding different features in it and finally generating textual description for that features [3][6] As limiting scope of image processing here we focus on face appearing in news image. Face recognition in basically considered as three step task as shown in fig. 2. First of all face is detected from given image which we have to recognize. Hemant Singh Mittal et. al. presents scheme based on Principal component analysis **(**PCA) and Neural network for person detection in image [7]. Viola et. al. introduces method for quickly and accurately detecting face in given image by generating classifier based on haar like features [22]. From detected face features are find out which are then further use for finding similar feature in face database. There are different face recognition approaches supervised and unsupervised like PCA, Linear discriminant analysis (LDA), SVM, HMM, Bayesian etc. Aleix et. al. discusses performance of PCA and an LDA algorithm based on different image sets and features and also states their behavior [1]. This paper in organized as, section 3 discusses about system architecture and overview of its modules. Section 4 discusses about experimental results of name entity recognition, face detection and face recognition and finally conclusion and future work in section 5.

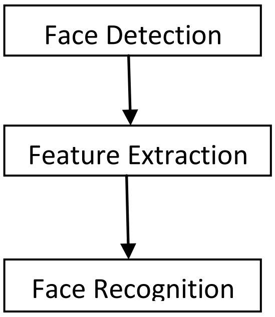


Figure 2 Steps in Face Recognition

3. SYSTEM OVERVIEW AND ARCHITECTURE

Caption helps in understanding image and related document in shortly. Caption generation may be extractive or abstractive i.e. may be by extracting words or sentences in document as it is or generating own sentence without direct use of document content.

3.1 SCOPE

There are many news sites over Internet across worlds which are having thousands of its readers. Our main purpose here is to generate caption for images which are embedded in news articles. For that we are going to focus on text appearing in news and image in that article. In case of text here we are going to find out name of persons appearing in article, so we limit scope to Indian names only. We develop NER models for Indian names by training it on Indian news only. On other hand in case of image processing here we are going to just find out who's in picture? i.e Face recognition. For it in depth limits scope to only bollywood celebrities. We are going to develop face database of only bollywood celebrities. So finally scope of captions generation will be limited to only bollywood news with bollywood celebrities face images. Proposed system architecture for caption generation for news images is shown in fig. 3. Different modules used in it are as follows.

3.2 TERM EXTRACTION

News document contain both image and text which describe news. Text processing is applied on text part, initially pre-processing is done. Sentences are tokenized; there are different stopwords which basically does not have any meaning and are mostly removed from tokens which gives list of meaningful and useful words which helps in finding important keywords. Stemming is applied on remaining words which process each word and removes morphological and inflexional endings from words in English. There are different stemming algorithms like Porter, Lovins stemming [12]. Weights of every term is find out and from that high weighted terms are selected which helps in understanding of news. As there are different methods of term weighting as we seen in literature survey. These words then used for forming keyword list which will be used for caption generation. In text categorization term frequency plays important role. Each document represented as d=(t1,w1; t2, w2;....;tn,wn), where ti is a term, wi is the weight of the ti in the document d term frequency find out number of times term occur in document as,

$$W(d,t) = TF(d,t) \text{-------} (1)$$

When there are number of documents then IDF factor is used which determines term is common or rear across all document.

$$IDF(t) = \log (N/n) \text{--------} (2)$$

Where N denotes total number of documents and n shows number of document in which term t occurs.

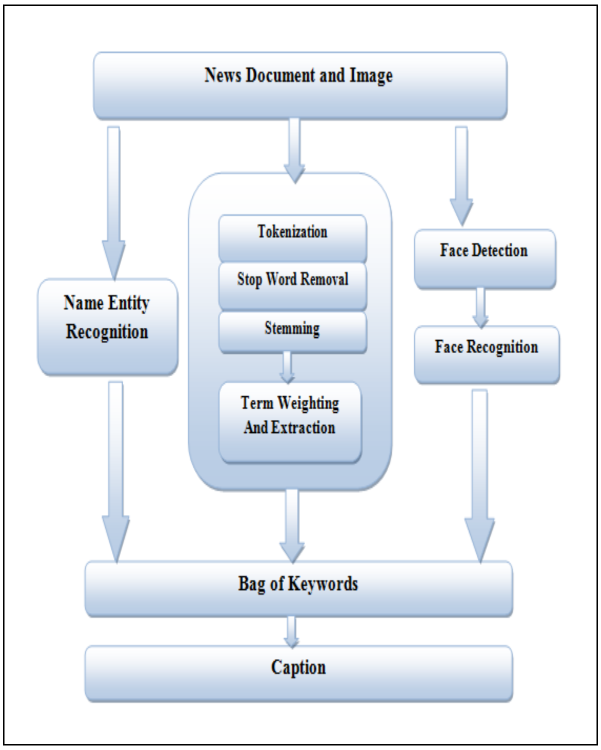


Figure 3 System Architecture

3.3 NAMED ENTITY RECOGNITION

With the enormous quantity of textual data available via the Internet and other electronic sources it is no longer feasible for human beings to process this data to identify useful information. There are number of things like name of person, location name, organization, date, money etc. which indicates that document is saying something related to that person or discussing about that organization. All such words with combination will helps in describing whole document or news in fewer words. It is very time consuming and tough task for human to extract such words. In information extraction domain named entity recognition is task which locates and classify word into predefines class like person name, location, organization, money, date, time etc. In case of news article which discuss about particular person or event, image in that article mostly contain face of that person. So it is important that name of that person must come in caption as it helps in retrieval of image when user query for a that person. Stanford provides three different models of NER as 3 class, 4 class, 7 class models, which are trained on different databases. 3 class model locates Location, Person, Organization in given text document while 4 class model extracts Location, Person, Organization, Misc and 7 class model focuses on Time, Location, Organization, Person, Money, Percent, Date in document [8]. On the other hand Apache provides OpenNLP library for natural language processing which is machine learning based toolkit [12]. General architecture for text engineering (GATE) also provides language processing tools, with name entity recognizer as gazetteer [11]. It consist of list of different entities like person name, organization, location and then this lists are used for finding occurrence of these names in text.

3.4 IMAGE PROCESSING

Caption must describe image features, its content as it decides image retrieval accuracy. Limiting scope of image processing we focus only on faces occurring in image. Face recognition is very critical and important task in a security section. It is somewhat easy in case of frontal face recognition but goes tough in case of partial faces which founds in uncontrolled scenarios like recognition in video surveillance frames and images captured by handheld devices. Shengcai Liao et. al. introduces partial face recognition technique which does not require any face alignment points or coordinates of any facial component in an image by developing multi-keypoint descriptors based alignment free face representation method [25]. Face detection is initial step in face recognition as detected face is then matched with faces available in database. Face detection accuracy also affects accuracy of face recognition.

3.4.1 FACE DETECTION

Face detection is technology which finds out face region in given image. There are lots of areas where face detection plays vital role like face recognition, security section in which real time face detection is done. In detection human face is get located in given image using database of different face image. Accuracy of face detection depends on different face images stored in face database. It is nothing but task of object classification, in which particular features are find out first for every class while training and at testing, objects are get classified as per their features class. Now a day's digital cameras also providing this feature which automatically locate face, which helps user for capturing image in less time and accurate. In case of face recognition face is recognized, and find out who is this person with the help of large database of face images of every person. For that first face is find out in given image and then that face is matched with database i.e. face detection is initial step in face recognition. There are many challenges while face detection like face position, facial expression, occlusion, image orientation etc. There are different methods which are categorized into four basic types as knowledge based methods, feature based methods, template based methods and appearance based. In knowledge based methods different rules are defined with the help of human knowledge of face and these rules then find out relationship between facial features. In feature invariant approach, it finds out structural features which remains in existence even if some changes occur like lightening effect, change in position. Using these feature face is located in this approach which is robust to such changes. In template matching methods standard templates are stored in database which are used for face detection by matching these templates with test face. Appearance based methods are like classifiers in which models are learned first on some training datasets and then these models are used for detecting face in test image. In this approach training dataset defines accuracy of model's face detection, if any new face or any change in test face occurs which is not in training set, then it is unable to identify that one [18]. Haar classifier helps in rapidly detecting objects from given image like human face, ear, nose, pedestrian etc.

A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window [22]. The presence of a Haar feature is determined by subtracting the average dark-region pixel value from the average light-region pixel value. If the difference is above a threshold that feature is said to be present. As there are hundreds of Haar features are presents at every image, so Viola and Jones simplifies it by introducing Integral Image technique, which adds small units together. For selecting specific Haar feature and to set threshold levels, AdaBoost machine learning method in used. It combines many weak classifiers and form a strong classifier from them. Numbers of AdaBoost classifiers are combined as filter chain which helps for efficiently classifying image region. The acceptance threshold at each level is set low enough to pass all face examples in the training set. The filters at each level are trained to classify training images that passed all previous stages. While testing, if any one of these filters fails to pass an image region, that region is immediately classified as "Not Face." When a filter passes an image region, it goes to the next filter in the chain. Image regions that pass through all filters in the chain are classified as "Face". Advantage of Haar like features over other features is its speed of calculation; it requires minimum time for face detection. Because of its Integral image concept Haar like feature of any size can be calculated in constant time.

3.4.2 FACE RECOGNITION

Detected face is then transformed to faces available in database for recognition. Using face recognition we can find out name for face. There are three main and commonly used methods for face recognition as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Linear Discriminant Analysis (LDA). PCA [16] finds a set of the illustrative projection vectors such that the projected samples retain the most information about original samples. Whereas ICA captures both second and higher-order statistics and projects the input data onto the basis vectors that are as statistically independent as possible. LDA [4] focuses on class similarities and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter.

3.4.2.1 PCA

PCA is an unsupervised face recognition technique. PCA manages the entire data for the principal components analysis without taking into consideration the fundamental class structure. The purpose of PCA is to reduce the large dimensionality of the data space to the smaller intrinsic dimensionality of feature space, which is needed to describe the data economically. In the training phase, feature vectors are extracted for each image in the training set. For it, image is first convert into pixel vector by concatenating each of rows to a single row. PCA works as dimensionality reduction technique which transforms pixel vector into feature vector which reduces dimensionality. For each training image these feature vectors are calculated and stored. While recognition, test image is given of known person. For that test image also feature vector is calculated as like done in training using PCA. This feature vector is matched with all training feature vectors and similarity between them is calculated. Similarity measurement is done using Euclidean distance method. Most similar feature vector in training set is resulted as recognized person [16].

3.4.2.2 LDA

LDA is supervised technique used for face recognition. It tries to find a basis for projection that minimizes intra class variation but preserve inter class variation. It group images of same class and separate out images of different class. Training set in LDA must contain images with diverse facial features. Each image is then represented in form of face vector. By storing same persons images in same place and different at different at training, cluster separation analysis is performed on this feature space. Using labels of classes and instances, within class and between class scatter matrices are computed. When face images are projected into discriminant vector, within class matrix shows how face images are distributed closely within class and between class matrix shows how classes are separated from each other. This discriminant vector tries to minimize within class variation and maximize between class variations. We perform some experiments on some standard face database, discussed their results, also created own database of bollywood celebrities and find out how they react to PCA and LDA, all this discussed in section experimental results.

3.5 CAPTION GENERATION

Term weighting gives most frequently word occur in news document, on other hand NER find out name entities like name of person, location, date. Here we take only name of person from news. List of keywords is formed from these words, also face recognition provides name for faces appearing in image. Bag of words which is generated from above keywords will be used for caption generation. As it contains all important words in news article, caption generated from it will explain whole news and fit well to an image. In text summarization domain lot of research is done and also going on. In it, sentence generation or realization also got a big interest. Different grammar rules must be applied while generating meaningful sentence from bag of words. Claire et. al. [5] presents feature based regular tree grammar approach for sentence realization, also there are some machine learning approaches used for headline generation for news [24]. Headline is also a few word sentences which in short describe whole news. As shown in our proposed methodology, from bag of words we are going to generate particular length statement as caption for image.

4. EXPERIMENTAL RESULTS

We carried out some experiments of Name Entity Recognizer on some test data consist of bollywood news, discussed results of Standard models provided by OpenNLP and Stanford, also trained own model and its results are compared with existing systems. For face detection, we performed experiment using haar classifier, find out its performance, results on test data.

In face recognition we first find out results of PCA and LDA on standard face databases, then on own bollywood face database. Also find out their performance by adding Gaussian noise to testing image.

4.1 EXPERIMENT 1: NAMED ENTITY RECOGNITION

We are going to find out names of celebrities in news and use this names while caption generation. There are different models i.e. NER models are provided by Stanford University, OpenNLPs NER model, GATE’s model etc. This models works well for European names, but it not gives that much results for Indian names. So here we are going to build OpenNLP’s name finder model by training it on own training dataset for Indian names.

Training Data: Training dataset is formed by collecting different types of news from different Indian news websites like NDTV, ZEE News, bollywood news sites, google news etc. We have done testing on different sets of training datasets by varying number of statements in it.

Testing dataset: So here we have made own dataset for testing. Stanford and OpenNLP NER models results are find out on this testing dataset. Total 50 statements are collected which discussed about bollywood celebrity “Shah Rukh Khan” from NDTV news site, dated in between 25 to 30 April 2014.

Total number of statements: 50

Total number of words: 775

Actual number of times name appear in testing document: 52

Table 2 Result of Different Training Dataset on Test Data

Sr No	Training Models (No. of Sentences)	Precision	Recall	F Measure
1	5000	0.5172	0.5769	0.5452
2	10,000	0.5517	0.6153	0.5816
3	15,000	0.5862	0.6538	0.6180
4	20,000	0.6551	0.7307	0.6907

As shown in above table 2, recall value for model 1 is less than 2 i.e. it indicates that as number of statements in training increases, model get trained well on different formats of sentences and it will be able to find out names in test data and correct predictions increases accordingly. For better accuracy training dataset at least contain 15,000 sentences with variations in format.

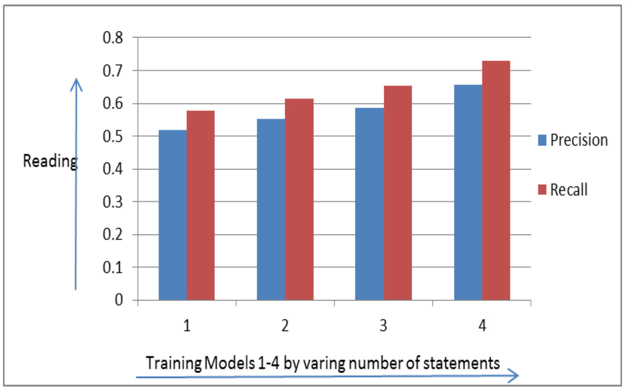


Figure 4 Graph of Precision and Recall of Own training model with varying number of entries

Above graph in fig.4 shows performance of four models over precision and recall values after testing it on test dataset.

We applied Stanford’s NER, OpenNLP’s default name finder model and our own trained OpenNLP name finder model on same testing dataset mentioned above.

Table 3 Testing results of Stanford NER, OpenNLP and own trained model

Sr No	Models	Precision	Recall	F Measure
1	OpenNLP default name finder	0.1923	0.0961	0.1276
2	StanfordNER	0.5370	0.5576	0.5473
3	Own Trained OpenNLP NER model	0.5689	0.6346	0.5999

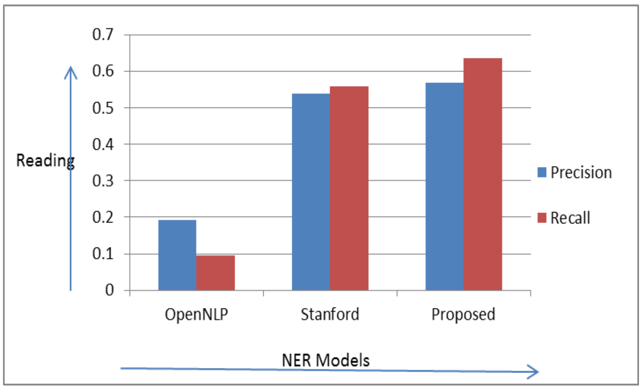


Figure 5 Graph of Precision and Recall of OpenNLP, Stanford and Own trained NER model

Above table 3 and graph in fig.5 shows that OpenNLP’s model performs poor than Stanford’s. Our trained NER model performs well than OpenNLP’s default name finder and gives results little more accurate than Stanford’s model. So in our proposed approach we are using own trained OpenNLP’s name finder model for finding names in news.

4.2 EXPERIMENT 2: FACE DETECTION

As image in article may contain more than one face in it. So first faces are found out and then it is given for recognition.

For face detection we have tested Haar classifier on test dataset consist of images of bollywood celebrities taken from NDTV news site. Some images contains more than on face.

Total number of test images: 50

Actual number of faces in all images: 97

Table 4 Haar classifier results

Correctly Detected	False Detection	Accuracy	False rate
90	15	92.78%	15.47%

As shown in table 4, rate of correct face detection of haar classifier is 92.78% which is well as false rate is also less. Fig.6 shows detected faces in news images, as it shows sometimes it detects false regions also as face, as seen in last image on right corner.

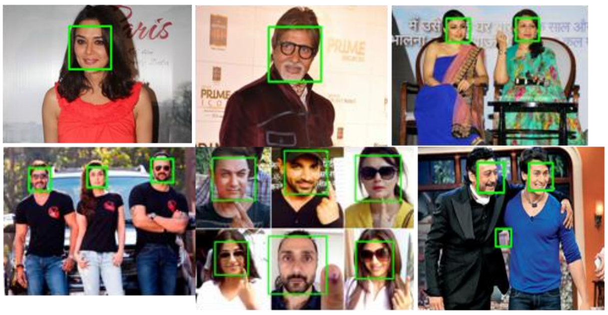


Figure 6 Detected regions as faces by Haar classifier

4.3 EXPERIMENT 3: FACE RECOGNITION

For finding name for person appearing in image we did experiments on some standard datasets, find out results of PCA and LDA algorithms, and then find out results on own dataset. Table 5 shows performance of PCA and LDA over two standard datasets.

Table 5 PCA and LDA results on Standard database

Sr No	Database	Images	PCA	LDA
1	AT&T ORL Face database [10]	400	39.5%	94.75%
2	IIT Kanpur Face database [9]	671	24.28%	81.23%

Accuracy of PCA is very low in both cases, AT&T and IIT Kanpur face dataset as 39.5% and 24.28% respectively. Whereas LDA gives better results as 94.75% on AT&T and 81.23% on IIT Kanpur face database.

Now we going to form own face dataset for face recognition.

Training Dataset:

As mentioned in section 3 our scope is limited to bollywood news and its celebrities, we made own database of faces of 10 bollywood celebrities. The files are stored in PGM format. The size of each image is 92x112 pixels. Images are organized in 10 directories, which have names of the form sX, where X indicates the subject number. Each of these directories, with number images of that subject, which have names of the form Y.pgm, where Y is the image number for that subject.

Testing dataset: We have made own dataset for testing. It consists of 50 face images of bollywood actor Amitabh Bachchan taken from www.google.com search results for query “Amitabh Bachchan”. All images are resized to 92x112 and converted to PGM format. It contains face with different face poses, with and without glasses.

We have performed testing on own dataset made for testing by varying number of subjects in training set per class as 5, 10, 15, 20, 25, 30 for checking algorithms performance. Following table shows results of PCA and LDA on testing dataset. Table 6 shows that as number of subjects per class increases performance of LDA goes on increasing as it does. On other hand performance of PCA reduces somewhat as subjects per class increases.

Table 6 Performance of PCA and LDA over varying dataset size

Sr No	Subjects Per Class	PCA Accuracy	LDA Accuracy
1	5	12%	16%
2	10	10%	18%
3	15	14%	40%
4	20	16%	34%
5	25	10%	46%
6	30	10%	54%

As we see in theory that LDA focuses on between class scatter than within class scatter, as number of different subjects per class increases result of LDA goes on increases. So here model 6 i.e. when subjects per class are 30, LDA gives better results than other.

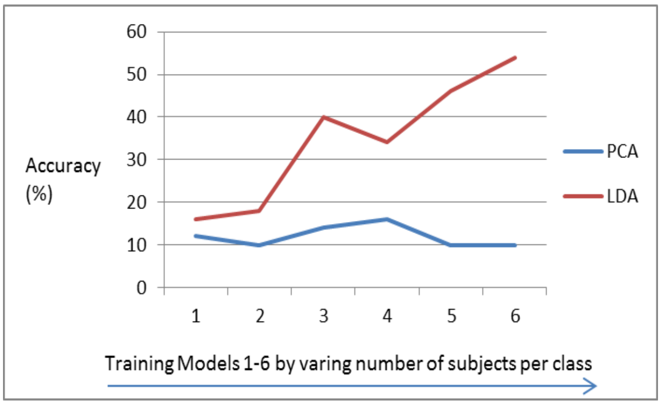


Figure 7 Graph of Accuracy (%) of PCA and LDA over different training dataset varying as subjects per class

Above graph in fig.7 shows performance of LDA goes on better side as PCA gets degrades as number of subjects’ changes per class.

As discussed above, dataset with 30 subjects per class gives better results than other, now we used same for further testing. Now we add noise in test images (testing dataset same as stated above) and check performance of PCA and LDA. Here we add Gaussian noise to each image step by step. Following table shows results of PCA and LDA on noise,

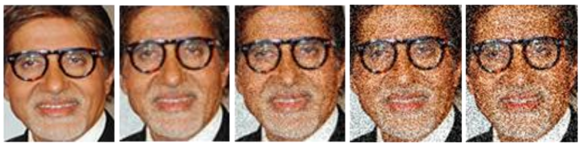


Figure 8 Test images of Amithabh Bachchan with increasing rate of Gaussian noise

Table 7 Performance of PCA and LDA over Gaussian noise

Sr No	Noise Ratio	PCA Accuracy	LDA Accuracy
1	10%	14%	54%
2	20%	12%	48%
3	30%	8%	56%
4	40%	14%	52%

Fig.8 shows sample test image of Amitabh Bachchan, first one is without noise and next four are with addition on noise in increasing order. Table 7 shows results of PCA and LDA over Gaussian noise. It shows that LDA recognize face even if noise ratio increases but variation in flow i.e. not always good. But it performs well over PCA. Below graph in fig.9 shows variation in accuracy of PCA and LDA when noise is added to images.

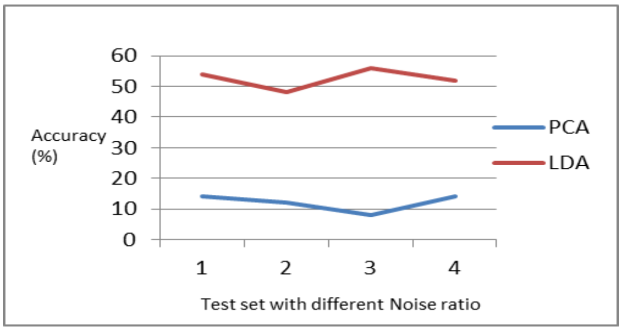


Figure 9 Graph of Accuracy(%) versus different test sets with noise variation for PCA and LDA

In our proposed methodology NER, face detection and face recognition are main modules. As we have seen performance and results of NER, face detection and face recognition of different methods, initially for name entity recognition we are using our own trained OpenNLP's NER model. Then using haar classifier we are going to find out face in given image and as LDA performs well over PCA, detected face is given as an input to LDA for face recognition, which gives name for that face if it exists in database. Using weighted terms and terms given by above modules, bag of keyword will be formed which helps in generating caption.

5. CONCLUSIONS AND FUTURE SCOPE

In this paper we have presented new approach for caption generation for images which are embedded in news articles. We mainly focus on important terms in news, named entities and face occurring in image. In this paper we have presented and discussed experiments which we have carried on NER and face recognition. Number of conclusion which are drawn from these experiments are as follows:

- Stanford's NER and OpenNLP's name finder model gives poor results for Indian names.
- We train OpenNLP name finder model for Indian names and it performs well over default models.
- Haar classifier detects faces accurately in short time and false rate is also less.
- LDA performs better over PCA.
- Performance of LDA increases as subjects per class increases.
- LDA does well over PCA when Gaussian noise is added in image.

A list of keyword will be formed from these three modules i.e. high weighted terms, person names given by NER and name given by face recognition. From this list short caption is formed. In future we can develop OpenNLP name finder model more accurate by adding more number of entries with variation in statement and person names. Also by varying training image i.e. considering faces with different features like lightening effect, noise, face angle we can improve accuracy of face recognition algorithm.

Approach presented in this paper for caption generation helps in generating automatic caption for images which are embedded in news document and as all useful and important keywords are focused caption generated from them also makes it content specific. Name entity recognizer and face recognizer together will help in increasing caption strength. This content specific caption helps search engines while retrieving specific images and also helps for improves image retrieval accuracy.

6. ACKNOWLEDGMENTS

This research is under study with the support of the teachers of Pune Institute of Computer Technology, University of Pune. It would be a pleasure to thank Emmanuel M. for his valuable suggestion and direction contribution that helps us in many ways to complete this research. Also thankful of IIT Kanpur and AT&T for providing open access for face database..

7. REFERENCES

[1] Aleix M. MartõÂnez, Avinash C. Kak. 2001. "PCA versus LDA". IEEE Transactions On Pattern Analysis And Machine Intelligence. Vol. 23. no.2. pp. 228-233.

[2] Allan Hanbury. 2008. "A survey of methods for image annotation". Journal of Visual Languages and Computing Elsevier. Vol. 19, Issue 5. pp. 617–627.

[3] Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee, and Song-Chun Zhu. 2010. "I2T: Image Parsing to Text Description". Proceedings of the IEEE. Vol. 98. no.8.

[4] Bhattacharyya, Suman Kumar, and Kumar Rahul. 2013. "Face Recognition By Linear Discriminant Analysis". International Journal of Communication Network Security 2. Vol.2. Issue 2. pp. 2231-1882.

[5] Claire Gardent , Benjamin Got T E Sman, Laura Perez-Beltrachini. 2011. "Using Regular Tree Grammars to enhance Sentence Realisation". Published in "Natural Language Engineering 17. pp. 185-201.

[6] Girish Kulkarni, Visruth Premraj, Sagnik Dhar, Siming Li, Yejin Choi, Alexander C Berg, Tamara L Berg. 2011. "Baby Talk: Understanding and Generating Image Descriptions". IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1601 – 1608.

[7] Hemant Singh Mittal, Harpreet Kaur. 2013. "Face Recognition Using PCA & Neural Network". International Journal of Emerging Science and Engineering (IJESE). Vol.1. Issue 6. pp. 71-75.

[8] http://nlp.stanford.edu/software/CRF-ER.shtml, 15/05/2014

[9] http://vis-www.cs.umass.edu/~vidit/AI/dbase.html, 27/04/2014

[10] http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html, 25/04/2014

[11] https://gate.ac.uk/, 10/052014

[12] https://opennlp.apache.org/, 20/05/2014

[13] Ilia Smirnov. (2008). "Overview of Stemming Algorithms". Mechanical Translation.

[14] J. Jeon, V. Lavrenko and R. Manmatha. 2003. "Automatic Image Annotation and Retrieval using CrossMedia Relevance Models". Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval.

[15] Jia-Yu Pan, Hyung-Jeong Yang, Pinar Duygulu and Christos Faloutsos. 2004. "Automatic Image Captioning". IEEE International Conference on

Multimedia and Expo, ICME '0., Vol. 3. pp. 1987-1990.

[16] Kim, Kyungnam. 1996. "Face recognition using principle component analysis". In International Conference on Computer Vision and Pattern Recognition. pp. 586-591.

[17] L. Ferres, A. Parush, S. Roberts, and G. Lindgaard. 2006. "Helping People with Visual Impairments Gain Access to Graphical Information through Natural Language: The igraph System". Proc. 11th Int'l Conference On Computers Helping People with Special Needs. pp. 1122-1130.

[18] M.-H.Yan, D.Kriegman, and N.Ahuja. 2002. "Detecting faces in images: A survey". IEEE Transaction on Pattern Analysis and Machine Intelligence 24. no. 1. pp. 34-58.

[19] Man Lan, Chew Lim Tan, Jian Su. 2009. "Supervised and Traditional Term Weighting Methods for Automatic Text Categorization". IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.31. Issue 4. pp. 21-735.

[20] Michele Banko , Vibhu O. Mittal , Michael J. Witbrock. 2000. "Headline Generation Based on Statistical Translation". Proceedings of the 38th Annual Meeting on Association for Computational Linguistics. pp.318-325.

[21] Mr. Saurabh M Khatri, Emmanuel M., Dr. Ramesh Babu D. R.. 2013. "A Novel scheme for Term Weighting in Text Categorization: Positive Impact factor". IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 2292-2297.

[22] Phillip Ian Wilson, Dr. John Fernandez. 2006. "Facial Feature Detection Using Haar Classifiers". Journal of Computing Sciences in Colleges archive, Vol. 21. Issue 4. pp 127-133.

[23] Roi Blanco, Christina Lioma. 2012. "Graph-based term weighting for information retrieval". Journal Information Retrieval. Vol. 15. Issue 1. pp. 54-92.

[24] Ruichao Wang, John Dunnion, Joe Carthy. 2005. "Machine Learning Approach To Augmenting News Headline Generation". In Proceedings of the International Joint Conference on Natural Language Processing.

[25] Shengcai Liao, Anil K. Jain, Fello and Stan Z. Li. 2013. "Partial Face Recognition: Alignment-Free Approach". IEEE Transaction on Pattern Analysis and Machine Intelligence. Vol.35. Issue 5. pp. 1193-1205.

[26] V. Lavrenko, R. Manmatha, and J. Jeon. 2003. "A Model for Learning the Semantics of Pictures". Proc. 16th Conf. On Advances in Neural Information Processing Systems.

[27] Yansong Feng, Member and Mirella Lapata. 2013. "Automatic Caption Generation for News Images". IEEE Transactions on Pattern Analysis and Machine Intelligence.Vol.35. Issue 4. pp.797-812

[28] Zenghai Chen, Hong Fu, Zheru Chi and David Dagan Feng. 2012. "An Adaptive Recognition Model for Image Annotation", IEEE Transactions on Systems, Man, and Cybernetic Part C: Applications and Reviews. Vol.42. Issue 6. pp.1120-1127.