



Contents lists available <http://www.kinnaird.edu.pk/>

Journal of Natural and Applied Sciences Pakistan

Journal homepage: <http://jnasp.kinnaird.edu.pk/>



VIDEO ANNOTATION FRAMEWORK FOR NEWS HEADLINES USING DEEP VISUAL AND TEXT INFORMATION

Luqman Izhar^{1*}, Shazia Arshad¹, Urooj Saeed¹, Muhammad Usman Ghani Khan¹

¹Department, of Computer Science, UET, Lahore

Article Info

*Corresponding Author

Email: izharluqman@gmail.com

Abstract

TV News channels are broadcasting news bulletins every hour. News bulletins data set is growing each hour for video tagging and retrieval. News headlines bulletins are a combination of different news annotations. International news channels media houses rely on local news sources to track a news of a specific region. News team research group digs in manually to monitor trends of local news sources to determine if a news of a specific region can create a wave as international news. To select a specific news annotation from local news sources requires consideration of multiple real time factors like coverage, rating and impact. Proposed Research model solves this real time problem by using deep learning method .Framework proposed classifies and retrieves a video annotation based on news indexing, priority and quality. Data set of 250 news bulletins was gathered from top seven news channels for top ten global categories like politics, terrorism, finance, weather & entertainment etc. Framework classifies the news annotation retrieval and tagging using 2D CNN with validation accuracy of 98.14% by calculating and considering real time parameters news indexing, priority and quality across cross bulletins contributing to novelty of the research.

Keywords

Annotation, Headlines, Indexing, News, Priority, Video, 2DCNN

1. Introduction

Number of News channels has almost doubled in last few years in every part of world. With the increase in number of news channels it is now difficult to view and track particular news. News channels are publishing news bulletins every hour. Prime time news bulletins are on average forty to sixty minutes long. To watch complete news bulletin is time consuming for viewers. Intentional media houses and their news sources track local news channels to monitor news that has a potential of making it to International news market like Elections, Terrorism etc. Along with international news channels Local/National media houses are themselves tracking local news bulletins to organize their hourly news headlines. To track all news sources consumes a lot of manual efforts. Automation of the whole process using deep learning principles can help research team of media houses to select news for their news bulletins. A framework that can intelligently predict and retrieve a video based upon on real time factors seems to solve this challenge.

Computer vision based image processing models have ability to retrieve video channels using textual information embedded in image, a few techniques uses voice data a few techniques uses contextual data and most techniques use face detection method. However there is lack of technique that can take into account real time factors (News index in bulletin, News priority across bulletin, News video Quality) for

video categorization. A model that also considers not only global features but also local features out of the input channel and can retrieve accordingly

Models of Computer vision and AI have robust solution for image processing data sets classification. While with Video categorization techniques are still evolving

The research work proposes a unique contextual solution for Annotations based video retrieval. The framework first segments a particular news clip using scene boundary detection method which separates different news from each other. After that Video classifier sorts the clips by extracting their deep visual and text information. Operations are performed on Frame repository to calculate indexing, priority and quality. The Framework outputs the probability of News Annotation. In Last, all news headlines videos are retrieved for Annotations which achieved max predicted probability output based on input image/video given by user.

2. Related Work

Researchers have used 2D- Convolution Neural Network for the training of models. That uses data sampled from 200 videos from News channels. Later, videos were converted into frames depending upon the eight proposed categories which defines the story Type. The proposed methodology achieves 91.7% accuracy as compared to AlexNet and VGG-16 architectures (Hassan, *et al.*, 2019).

Khan & Hassan, converted the frames of video into chunks based on category and extracted motion vector threshold by using both temporal and spatial information in 3DCNN. 3DCNN achieved more accurate results as compared to LSTM which is a 2D CNN based method. (Khan & Hassan., 2018)

Singh, proposed a framework that takes into account the social aspects of the news to predict the news popularity by using “feature selection”. Data set of tweets and news articles were used. Parameters like no of images, no of videos, no of Gif, no of hashtags, category mentions etc were taken in account using techniques SVM (Linear, RBF) ,RF and GBDT. Results showed that Random Forest and gradient boosting techniques performs well. (Singh, 2018)

Namous, used eleven different Data mining algorithms like Multi-layer perception , Bagging, AdaBoost,, Random Forest, Naïve Bayes, K nearest Neighbor , Logistic regression and SVM (linera, polynomial, RPF and sigmoid) on online news popularity dataset and calculated F measure. Research showed that both Random Forests and Multilayer Perception with four hidden layers and 0.1 learning rate obtained the best result with F-measure of about 65%. (Namous, *et al.*, 2018).

Li., proposes a two stage model based on Content feature. It first extracts the global elements from news and classify them and then predicts the popularity in second stage selects local feature specific words. Methodology also proves that power of

prediction is higher for social and financial news as compared to sports and entertainment. The framework contributed enhancement in prediction power by boosting it 2.2% by using algorithm as SVM, KNN, Decision Tree, random forest, Ada Boost and Gradient Boosting for four news categories. (Li, Y. *et al.*, 2018)

Shah, achieved directional accuracy of 70.59% in trend prediction by developing sentimental analysis dictionary for checking the effect of news sentiments on stock market. Proposed methodology fetches the articles and performs a preprocessing operations. After that data was transformed into numerical vectors (unigrams, bigrams or trigrams) and was compared with data dictionary to determine the words polarity for calculation of scores. (Shah. *et al.*, 2018).

Malagrino, investigated stock market index by using “Bayesian Network structure” instead of using Neural Networks and SVM. The model proved to be simple and general that makes no assumptions and requires minimum input and out performs as a good expert systems compared to Neural Network based expert models by predicting next day closing directions. Mean accuracy with configuration that takes single index per continent was around 71% (with almost 78% top accuracy). (Malagrino, *et al.*, 2018)

Deshpande, used machine learning techniques to predict the popularity of news by using Random forest, AdaBoost and LPBoost on UCI machine learning dataset. Adaptive boosting turns out best results for

predictions with Accuracy of 69% having F measure of 73%. The process started with collection of data set and by reducing dimensions of data set by LDA- Linear discriminant Analysis and then applying ML algorithms. (Deshpande, 2017)

3. Pre Processing

Data set was collected from top seven leading news channels social web pages, official sites. Data set sampled was collected for same month covering same stories so indexing, priority can be determined. Data set tackled 10 different classes of Annotations. All news clips were converted into frames using Scene boundary detection. Corrupted and Pre story frames were also discarded in this process. Frames with news caster visuals, transition visuals and Channels visuals were sampled out

Table 1: Non Contextual Frames

Sr. No.	Non Contextual Frames
1	News channel Frames
2	Transition Frames
3	News Caster Frames

4. Inception Model

Inception model is Convolution neural network that uses CNN. The convolution is used to extract the features of the images by applying series of filters 1*1, 3*3, 5*5. Applying small filters results in extraction of local features and applying bigger filters results in extraction of global features out of the image. Since the framework needed to

work on both local and global features of images Inception model was used for this purpose

Inception model applies all convolution in parallel and select the best possible

- 1- Convolution layer helps in feature extraction
- 2- Pooling layer help in dimension reduction
- 3- Softmax layer outputs the probability distribution.

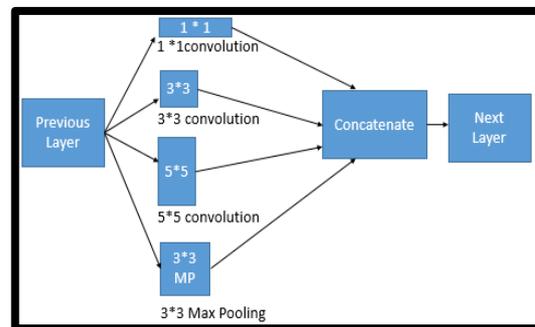


Figure: 1 Inception Model Layers

Inception model is pre trained model on which data set can be loaded and it can output results on 1000 objects categories. For this research work the association of 1000 object classes was not used and Ten Annotation classes was mapped to achieve the results

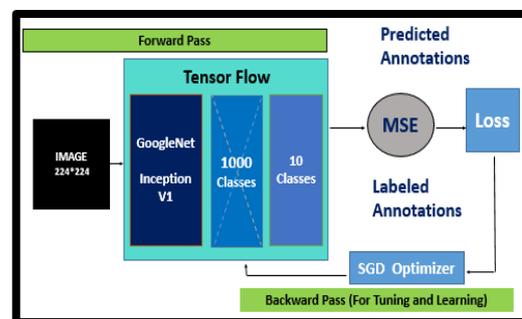


Figure: 2 Framework Architecture

5. Methodology

Dataset of 250 videos was sampled out into 10 different Annotations as per data dictionary mapping

Table: 2 Data Dictionary Showing Associations at Annotation Levels

Global Annotations In News Bulletins <i>Level - 0</i>	Global Annotations Breakdown <i>Level - 1</i>	Annotations in Local News <i>Level - 2</i>
		Press Release (a1)
Political	Government (a1)	Law (a1)
	Opposition (a2)	Border Security (a1)
		Protest (a2)
Finance	Corruption (a3)	Investigations-FIA (a3)
	Tourism (a4)	Tourist spots (a4)
Terrorism	Blasts	Bomb / Fire
Health	Pandemic (a5)	Dengue Fever (a5)
Weather	Rain/Cyclones	Corona
Entertainment	Film & Social Media	Floods
Sports	Worldcup (a6)	Twitter Tweets
	Olympics Local Event	Cricket/Soccer/ etc (a6)

Global annotation level 0 ‘GA’ in data dictionary can be divided into Level-N as follows

$$GA_{Level_k[i]} = GA_{Level_k+1[0]} \dots GA_{Level_k+1[n-1]}$$

The videos were sampled out based on specific story type using scene boundary detection.

$$Clip(a,b) = C_i [XOR] C_{i-1}$$

After that News priority, new indexing and quality is detected. Quality help in pulling out best Frames/ Videos depending on user query preferences.

For Clip ‘X’ having Annotation ‘A’. Following parameters were calculated

$$A \rightarrow Priority(X), Index(X), Quality(X)$$

Table.3 shows the parameter, expression to evaluate the parameter to understand the real time factors in News headlines. That can help in news retrieval and selection

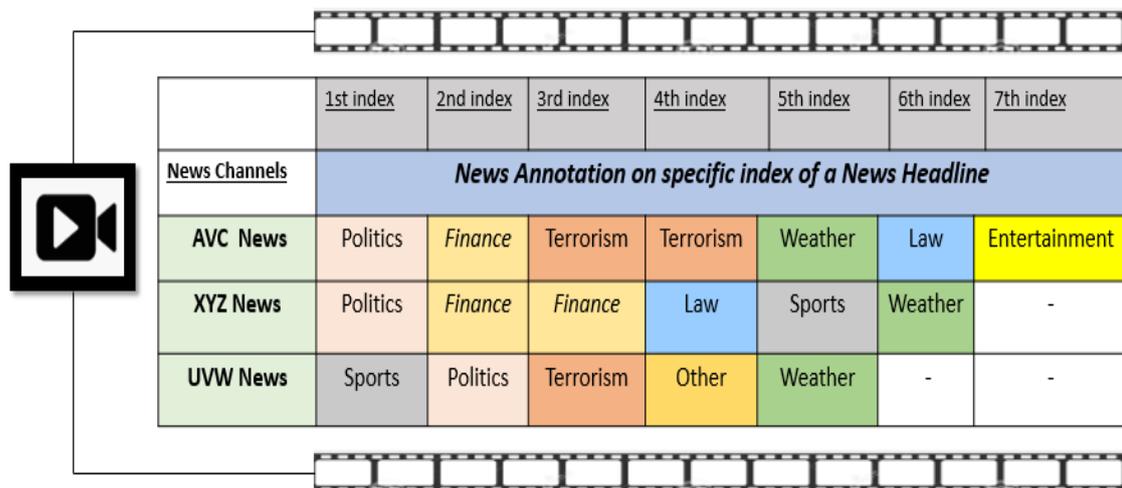


Figure: 3 Segmentation Of Annotation Across A Bulletins Of News Channel

Table: 3 Calculation of Priority, Indexing and Quality Attributes

Parameter	Expression	Details
Occurrence	Occurrence(X) = Integer 'N1'	N1 = Nth number of times a story appears in News bulletin
Duration	Duration(X) = Double 'N2'	N2 = Duration of each instance of clip
Priority P(X)	P(X) = Occurrence(X)+ Duration(X)	Combination of 'Occurrence + Duration'
Indexing I(X)	I(X) = Integer 'N3'	N3 = Ith number of sequence in which new exist in a specific bulletin
Quality Q(x)	Q(X) = FPS Count	Frames per second

5.1 Parameter Tuning

To achieve optimized results parameters were tuned via Stochastic Gradient Descent.

The batch size of 256 Dynamic learning rate was kept to 0.0004, drop value was scaled to 0.3 and Momentum was kept to 0.9. The framework was left to train to achieve the desired results.

The epoch level were portioned into 5 classes A: 0-1000, B:1001 -2000 , C:2001 -3000, D:3001-4000 and E: 4001 -5000

2048 Transfer values were converged to the scores of Annotation classes using softmax function.

$$f_i(\vec{a}) = \frac{e^{a_i}}{\sum_k e^{a_k}}$$

The single score of each class was the divided with sum of all scores to receive the probabilistic distribution

Probability of Annotation Ai = [Score of Ai] / [Sum of all scores of classes]

6. Dataset

Dataset available across internet was not satisfying the research requirements. So data set was gathered which is itself the novelty of the research. The data set was gathered from seven top news channels Dunya, GEO, ARY, EXPRESS, 92HD, SAMAA and Aap News. From Data set Ten Annotations were sampled out.



Figure: 4 Annotations of Specific News Frames in Bulletin

7. Results

The data set was segmented into training set and validation data set. The validation set was sampled out of training set. At epoch level 5000 the validation accuracy for trained annotation shows the accuracy of 98.14%. The validation/Testing accuracy is shown in bar graph Figure: 5

Framework Accuracy on different epoch level is listed below

Table: 4 Accuracy on Epoch Scale

Sr. no	Epoch	Accuracy
1.	1000	56.06%
2.	2000	65.06%
3.	3000	71.76%
4.	4000	89.06%
5.	5000	98.14%



Figure: 5 Validation Accuracy of Framework
Framework Accuracy for validation data set is shown in Figure.6

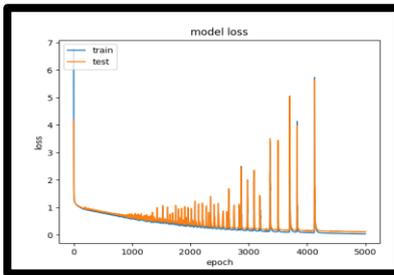


Figure: 6 Framework Accuracy

News Annotation	Input Frame	Retrieved Output
Entertainment		
Finance		
Tourism		
Security		

Figure:7 Input Provided Vs Predicted Output

8. Conclusion

Dataset was collected from seven different new channels. The data collected was segmented based on ten different Global Annotations based on the data dictionary formulated and co relations identified. The framework takes input the News clips converts the videos into frames

and applies 2DCNN to and output the probability of video Annotation by analyzing the news indexing, priority and quality. The framework proposed can contribute towards News Aggregation, Validation Analysis through cross examination, News popularity prediction in swarms. The research work also paves ways for future work in areas of Documentary generation and in News Rating Analysis

The research artifacts i-e data set collection, identification of famous news Annotations, Data dictionary and Framework are core novel contributions of this research. Out of which Data set and Framework will be shared with research community online

9. References

Deshpande, D. (2017). Prediction & Evaluation of Online News Popularity Using Machine Intelligence. 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA). doi:10.1109/iccubea.2017.8463790

Hassan, M. A., Saleem, S., Khan, M. Z., & Khan, M. U. G. (2019). Story Based Video Retrieval using Deep Visual and Textual Information. 2019 2nd International Conference on Communication, Computing and Digital Systems (C-CODE). doi:10.1109/c-code.2019.8680995

Khan MZ, Saleem S, Hassan MA, Khan MU. Learning Deep C3D Features For Soccer Video Event Detection. In2018 14th International Conference on Emerging Technologies (ICET) 2018 Nov 21 (pp. 1-6). IEEE

Li, Y., Peng, Q., Sun, Z., Fu, L., & Khokhar, S.-D. (2018). A Two-stage Prediction Method of News Popularity only using

- Content Features. 2018 13th World Congress on Intelligent Control and Automation (WCICA). doi:10.1109/wcica.2018.8630557
- Malagrino, L. S., Roman, N. T., & Monteiro, A. M. (2018). Forecasting stock market index daily direction: A Bayesian Network approach. *Expert Systems with Applications*, 105, 11–22
- Namous, F., Rodan, A., & Javed, Y. (2018). Online News Popularity Prediction. 2018 Fifth HCT Information Technology Trends (ITT). doi:10.1109/ctit.2018.8649529
- Shah, D., Isah, H., & Zulkernine, F. (2018). Predicting the Effects of News Sentiments on the Stock Market. 2018 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2018.8621884
- Singh, H. V. (2018). Predicting the Popularity of Online News using Social Features. 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT).