Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 7, Number 1 (2015), pp. 49-57

http://www.irphouse.com

A Survey on Multimedia Analytics using Big Data Analytics

T.V.M. Sairam and Dr. B. Thirumala Rao

M.Tech, Department of CSE, K L University, Vaddeswaram, Guntur
Professor, Department of CSE, K L University, Vaddeswaram, Guntur

Abstract

Multimedia is the only source to the world to present any concept in an apt format and can inject easily to any user to understand the concept easily. Basically multimedia provides information- rich data using which we can process and get information about anything easily. Multimedia includes images, videos and audio. There are several papers which explains the process of extracting natural language (NL) from that rich data, so that it will be easier to the user to search his/her required data easily. In this paper we are going to discuss about how the manual search can be replaced with automated search with multimedia analytics in big data, so that the user can use that processed data easily and can predict the future requirement. Topic –Oriented Multimedia Summarization (TOMS) is the novel approach in which there will be some challenges which are needed to resolve the low-level semantic gap between the objects.

Keywords: Multimedia Analytics, Natural Language, Information-rich, semantic gap, Topic-Oriented Multimedia summarization, Automated.

1. Introduction

Multimedia analytics is not a novel concepts through which we can summarize the available information- rich content for feature extraction. If we consider a general example in real time " If a student is searching for an online tutorial videos regarding his subject and if he enter the text as a key in the search bar then he will get all the videos which are not related and required also. In that huge list its difficult task to find out our required videos.". If this is the case if the data is summarized and annotation is applied by using some algorithms we can get the required data in an apt manner and it should be done automatically. So automated multimedia summarization have a challenge to prevent the low level semantic gap between the contents. Researchers are looking forward to start a new era in the field of multimedia analytics which allows the automated data processing and this research is also leading for remove the semantic gap as specified before. The main motto for this research is to work on different types of summarizations in multimedia and to provide a ease of use of the software around use in a simple manner for different types of organizations. First we need to find out the different categories that have the multimedia data more than any other. So if the survey worked out in all aspects to gather and process the unstructured data to structured format. There are different characteristics for different types of multimedia data. Let use first know about the characteristics of the video in multimedia analytics. 1) In video streaming if we consider, it consists of large amount of raw data which cannot be processed by normal processing methods, 2) The structure of the contents is small in size and can be varied, 3) The richness in the content is more than normal images. According to these characteristics we can confirm that the video is the information rich content than any other multimedia content. Because of these contents the [2] indexing and retrieval of the video content very difficult. In olden days i.e., few years back we used to use separate databases for the video and can be retrieved by a key word. But now its not a case, because of the availability of the larger content in the raw format. To avoid this we need to follow the process of content based information retrieval.

The applications for the content-based indexing and retrieval are as follows: 1) search can be according to the interest of the user, 2) remote instructions etc. we can detect the useful videos and trace the harmful videos and stop feeding them into our browser that comes under management of the web content. TREC and [3][4][5]TERCVid are the two sponsoring researches from NIST since 2001 using which users are enjoying content based retrieval of the videos. [1] TOMS (Topic –Oriented Multimedia Summarization) is the technique that is used to create text for the video in the streaming and not only for the video, even an audio clip can be divided into segments and can create meaningful semantic text for it for future easy search for the user. Let us discuss about each in detailed manner in further sections.

2. Related work

In this related work we need to discuss mainly on the work done by some researchers related to the multimedia summarization that is audio and video summarization.

2.1. TOMS (TOPIC – ORIENTED MULTIMEDIA SUMMARIZATION)

TOMS is used to create semantic text information from the current running stream of video. When we consider an example of a flash mob, Using TOMS system will recognize the most salient features in the video and if it is a good recounting system it will generate a passage in a semantic manner which describes the video about the people in the video, objects running in it and their characteristics like color, speech etc. the passage consists of the following data

A flash mob is going on with a group of people are dancing in between crowd,
Everyone cheering up, dancer were blue and black combination dresses and
Different tattoos on body.

Consider the above thing as a passage generated by the good recounting system and using this the user may extract his future information. Here comes a query that how the extracted information is used to predict the future or how it can be used for future use. If the recounting system extracts all the videos and arranged in the recommended search for the user so that he can view only his related data regarding flash mobs so that he can learn the new styles of presenting a mob. So that he can design a new style in the mob and can upload his video to the repository. Here comes another question that “what is the use of the recounting system and how we can now that the information given by this system is genuine?”. Here we need not to worry about anything. Because all the data gathered is based on semantics which will remove the love level semantic gap between the concepts. TOMS will summarize the video the most important information into text format which is belonged to certain area. Filtering is applied according to the recommendations of the video search. TOMS helps Optical Character Recognition(OCR), ASR, semantic audio concepts to present the systems results in natural and intuitive format.

2.1.1. Audio Summarization

By considering an audio for example a news streaming by extracting the salient words from it we need to summarize the audio and the text is created for ease of search and use[8]. In this ongoing research topic first a speech can be converted into text. To avoid the redundancy in the data we use the audio summarization categories. We need to calculate the amount of data is extracted at a given compression ratio[6][7].

Valenza et al[9] is one of the famous researcher who has given a method for measuring information retrieval and extraction techniques.

2.1.2. Video summarization

Large amount of multimedia data is available on internet in the form of video such as online classes, you tube videos, webinar’s, movies etc. but if a person is willing to watch news in online which are related to the current day he may get the apt information but with that some unwanted videos can also be available based on the recommended and the on the last search. On internet our result will also depend on the others search criteria. If a person watched video1 and next video2 and so on. For the other person who want video1 will also get video2 unwontedly which is useless data. So we need to abstract all the things and then we need to summarize all the data. If we consider two videos of same category but something is different in those two, then video summarization is applied by considering the salient words from that video. Even though the those two are same concept videos the content in the videos may vary.

2.1.3. Ranking video contents

Ranking is the new approach to categorize all the videos according to the user recommendation. Bipartite graph is one main example for representing the ranking for the videos. Rather than using the normal adhoc techniques we use this graphs technique for rankings. Rankings are of two categories. 1) visual concept based discrimination analysis 2) Re-ranking using ground truth. We shall discuss those in a short form by using small formulas.

1) *Visual concept based discrimination analytics:*

Let G=(V,C,E,W) be the bipartite graph between different videos and concepts. Each concepts is defined with a set of things. If we consider V, it is a set of training videos, C is the concepts in related to those videos, E represents Edge set, W represents the weights of each edge.

$$\begin{cases} f_{t+1}^{c} = \alpha \tilde{W}^{T} f_{t}^{v} + (1-\alpha) y^{c} \\ f_{t+1}^{v} = \alpha \tilde{W}^{T} f_{t+1}^{c} + (1-\alpha) y^{v} \end{cases},$$

y with v represents initial scores of video sets
for each event, the positive video node value is 1 and negative node value is 0
y with c represents the scores of concept sets
initial node value of the concept set is 0
f represents the updated score values of video and concept nodes
W represent normalized weights
D represents diagonal matrix. Sum of W in diagonal
α – Operation system (A propagation weight)
C+ positive V -> high scores
C+ negative V -> low scores

By considering each rank value of the concept and videos we apply the re-ranking and get the global ranking for the event.

2) *Re- ranking using ground concepts*

Here the importance is given to the human who will give the ranking to the training videos and the concepts of each event. Here is a score will be given to the videos and the concepts. Based on those values the importance of the video and concepts are determined.

$$Score_V(c) = 1 / (R(c)/65 + R_G(c))$$

C and R(c) describes about the visual concepts that are filtered.
R m (c) represents the human given ground truth value.
After gathering all the information we will perform the re-ranking to the videos and concepts.

2.2. Generic Module

Here we have pre-defined templates using which we design the passage for the video belongings. In this module we have 3 predefined templates as follows.

This is a <Topic_Name> event.

The video shows the event of <Topic_Name>.

This video is about <Topic_Name>.

2.3. Visual Concept Module

This module generates several sentences regarding the video. That is about the objects present in the video, characters, and explains about all the concepts in that. After giving the ranking to the video contents according to these passages generated manually or automated and final ranking is assigned. The syntaxes of those templates are as follows.

We saw <List_of_Visual_Concepts> in the video.

We <adv> saw <List_of_Visual_Concepts> in the video.

We can violate the adverb to include if the video object is identified by the key frames of the video generated in natural language.

2.4. Module for Text Concepts

Using the ranked automatic speech recognition techniques some automatic templates are generated and using this templates we include the rich texts such as objects and create semantic sentences for future research on the videos and audio files. Here we have a template which is used for describing the objects.

We <adv> heard the words <List_of_ASR Transcriptions> from the video.

3. Proved Experimental Paradigm

Using the TOMS system from the US Government [10] a researcher performed a small experiment using the pilot study. In this 20 videos are selected and 10 are manually decomposed into the text and remaining are by TOMS system. He selected university students to perform the experiment both manually and automated. Those videos are subdivided into event detection and video detection.

Using TOMS some passages are generated and using human creativity some are generated, each in case with human and TOMS. The table is as follows.

Table 1: Consists the events and video selection tasks

Generated by	Event Selection Task	Video Selection Task
TOMS	5 samples	5 samples
Human	5 samples	5 samples

Here it is a data set regarding the experiment. Some graphs are created according to the result. The graphs are as follows.

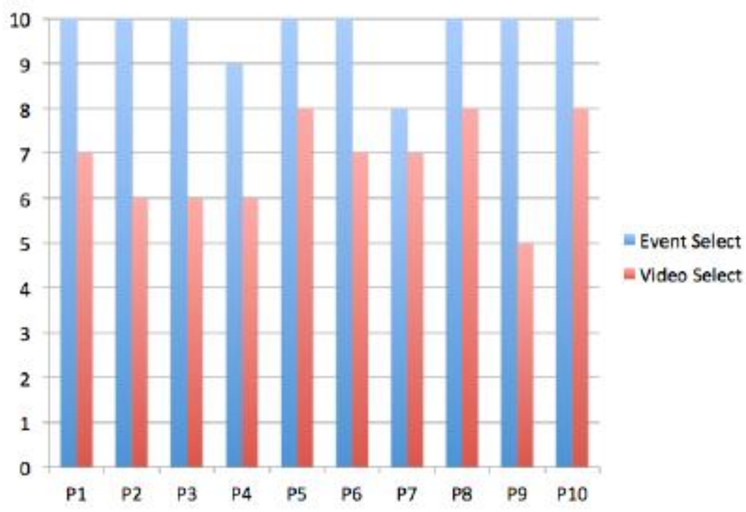

Figure 1: Ranking given to event and video by students of university

The above graph represents the pilot study regarding the events and the videos. The accuracy is measured in percentage for the future research is as follows.

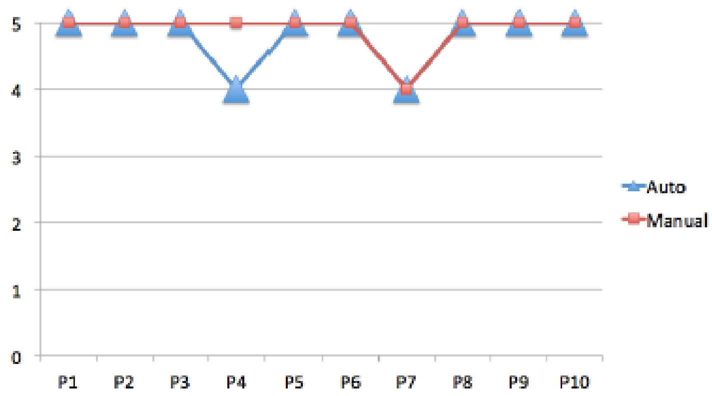

Figure 2: accuracy for events set by students manually and automated

This is the event selection task using the manual and automated using TOMS. The accuracy in these case is high than the video set.

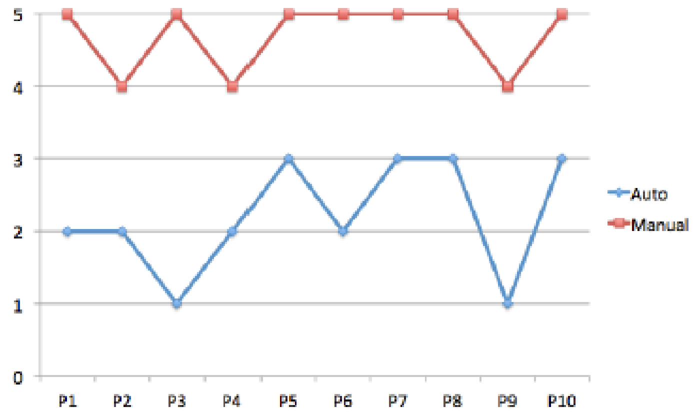

Figure 3: accuracy measure for the video events by students

This is for the manual and automated process for identifying the videos using key objects which are already identified.

4. Linear Regression using R- analytics.

R- analytics can be applied in the numerical data sets. Using TOMS we can create the numerical data sets so that those data sets can be subjected to the linear regression using the R- Analytics. R is the programming language and tool used for big data analytics and we can analyze the large set of the user generated data from various resources.

4.1. Linear regression results on "airquality"

A linear regression is conducted on airquality data set which is created already using the different resources. So here are the results for the airquality data set.

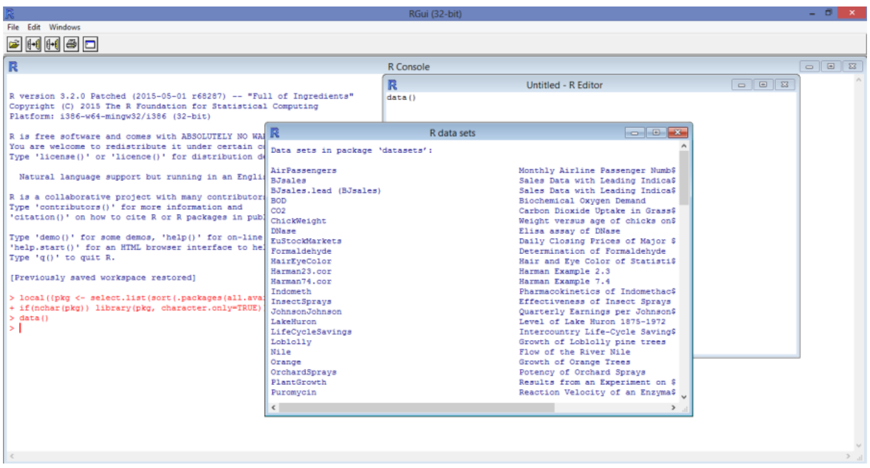

Figure 4: R- Tool console with the data sets generated for big data analytics

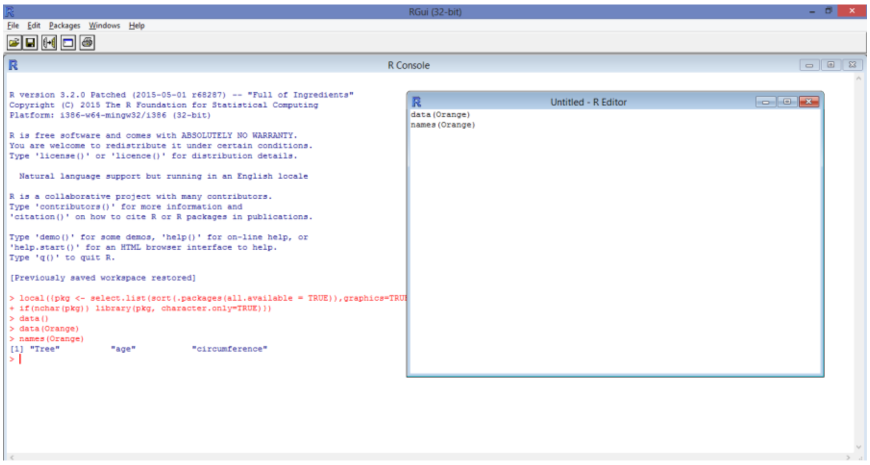

Figure 5: Getting the objects in individual data set

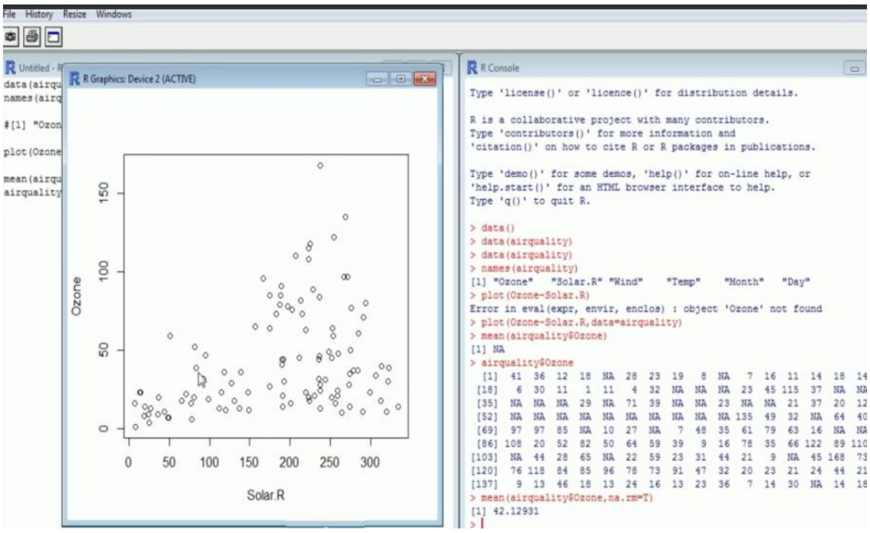

Figure 6: Graph generated using the events and values in the data set

Here we will compare Ozone and Solar.R to the linear regression. Some values are retrieved after analyzing that data. So the a graph is plotted using those values using the linear regression technique.

4.1. Process of getting the data set.

This process will undergone in few steps as follows.

Step-1: Frist the video or audio is considered and subjected to get the concepts from those events using the TOMS system which is an automated machine for transcripting.

Step-2: that datasets are subjected for the linear regression using the R- tool and R-programming language. To find out data set in R – tool use the following command. " data()" press Ctrl+R to compile in the console.

5. Conclusion And Future Work

The dataset created using the TOMS is processed using the R – Tool which is a big data analytics tool for the textual data produced using the datasets. In this paper I considered airquality dataset which is used for processing and getting the accurate information from the dataset. Right now we are considering the textual information from the video and audio. But research is still going on for the direct video and audio processing using the big data analytics.

References

1. D. Ding *et al.*, ``Beyond audio and video retrieval: Towards multimedia summarization," in *Proc. 2nd ACM Int. Conf. Multimedia Retr.*, 2012, pp. 2:1_2:8.
2. W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, ``A survey on visual content based video indexing and retrieval," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 41, no. 6, pp. 797_819, Nov. 2011.
3. P. Over, G. Awad, J. Fiscus, and A. F. Smeaton. (2010). "TRECVID 2009–Goals, tasks, data, evaluation mechanisms and metrics," [Online]. Available: http://www-nlpir.nist.gov/projects/ tvpubs/tv.pubs.org.html
4. A. F. Smeaton, P. Over, and A. R. Doherty, "Video shot boundary detection: Seven years of TRECVid activity," *Comput. Vis. Image Understanding*, vol. 114, no. 4, pp. 411–418, 2010.
5. A. F. Smeaton, P. Over, and W. Kraaij, "High-level feature detection from video in RECVID: A 5-year retrospective of achievements," *Multimedia Content Analysis: Theory and Applications* (Springer Series on Signals and Communication Technology) Berlin, Germany: Springer, 2009, pp. 151–174.
6. Yingbo Li, Bernardo Merialdo. Multi-video Summarization Based on AV-MMR. In *Proc. 2010 Int'l Workshop on Content-Based Multimedia Indexing*, 1-6.
7. Ani Nenkova. Summarization evaluation for text and speech: issues and approaches. *In Proc. INTERSPEECH 2006, USA.*
8. Peter Kolb. Experiments on the difference between semantic similarity and relatedness. *In Proceedings of the 17th Nordic Conference on Computational Linguistics - ODALIDA '09,Odense, Denmark, May 2009.*
9. Brian Langner and Alan Black, MOUNTAIN: A Translation- Based Approach to Natural Language Generation for Dialog Systems, *In Proc. of IWSDS 2009, Irsee, Germany.*
10. D. Ding *et al.*, ``Beyond audio and video retrieval: Towards multimedia summarization," in *Proc. 2nd ACM Int. Conf. Multimedia Retr.*, 2012, pp. 2:1_2:8.