Volume 5, Issue 11, November 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Web Multimedia Mining: Metadata Based Classification and Analysis of Web Multimedia Data

Siddu P. Algur, Basavaraj A. Goudannavar*
Department of Computer Science, Rani Channamma University,
Belagavi, Karnataka, India

Abstract—The astounding growth of multimedia data over the web, identifying and organizing the web multimedia data into different categories such as- 'Entertainment', 'News and Politics', 'Sports' etc is a complex and challenging task under the research theme – 'Web Multimedia Mining'. This work demonstrates identifying and organizing a domain specific web multimedia data into different categories using Data mining classification techniques based on metadata of considered web multimedia data. Using extracted metadata, web multimedia data are automatically classified into different categories by applying data mining classification algorithms such as Decision Tree (DT) and K-Nearest Neighbor classification (KNN) model. The classification results are compared and analyzed as a step in the process of knowledge discovery from web multimedia data.

Keywords—Web Multimedia Mining, Web Multimedia Metadata, Decision Tree Algorithm, k-Nearest Neighbor Algorithm.

I. INTRODUCTION

The incredible rapid growth of multimedia data on the Social media websites such as YouTube, Red Tube, and Face Book etc, automatic organizing of multimedia data into different classes is an emerging trend in the area of web multimedia research. When we look into the different classes of multimedia data, we found 3 different classes as discussed below:

Class 1: Combination of all the 4 basic components of multimedia data i.e. Image, audio, video and text, (Ex- Animation, News, Video lectures).
Class 2: Combination of any 3 basic components of multimedia data (Ex- Videos)
Class 3: Combination of any 2 basic components of multimedia data (Ex- PPTs)

The large numbers of all the 3 classes of multimedia data are growing day by day on the Internet. As multimedia data are increasing over the web, it is becoming difficult to identify and classify the multimedia data without knowing the content of it. In this experiment an attempt is made to classify Class 2 web multimedia data as a domain specific approach. Since, the video domain contains 3 basic components (i.e. audio, video and image), for experimental purpose, web videos have chosen as a Class 2 web multimedia data. The basic components of web videos will be separated for metadata extraction.

Many classification models/algorithms and data mining and machine learning tools are developed in recent years. In this work, using KNIME data mining tool [6], the web video metadata are extracted and classified based on available metadata of web multimedia-videos using Decision Tree and KNN classification algorithms. The classification results are compared and analyzed. The rest of the paper is organized as follows: The section 2 represents related works on the classification of web multimedia videos, section 3 represents proposed web multimedia video classification methodology, section 4 represents performance evaluation analysis of classification models, and finally section 5 represents conclusion and future work.

II. RELATED WORKS

The goal of classification is to build a set of data models that can accurately predict the class of different objects. Nowadays, numerous successful implementation of data classification in various applications using rough set theory are available. In this article they explain the rationale behind its creation, as well as the implications the dataset has for science, research, engineering, and development. Further present several new challenges in multimedia research that can now be expanded upon with our dataset. To summarize, our dataset has been curated to be comprehensive and representative of real world photography, expansive and expandable in coverage, free and legal to use, and intends to consolidate and supplant existing collections. They have highlighted its use for the next generation of computer vision, human mobility, and social computing research. Our dataset encourages the improvement and validation of research methods, reduces the effort of acquiring data, and stimulates innovation and potential new data uses [1].

In this paper, the semantic metadata required to describe the visual information content of videos are defined and classified into four distinct classes: Media Entities; Content Items; Events; and Supplementary Items, and three types of property tables are defined: Identity Tables; Spatio-Temporal Position Tables; and Event Tables, in which these metadata may be stored in a relational database [2]. The concepts and categories of semantic metadata defined in this paper described how metadata relating to the semantic information content of video may be classified for subsequent query by content. Specifically, intrinsic metadata from ancillary metadata, have differentiated between structural and semantic metadata within the intrinsic metadata class, and have described how the semantic metadata types should be organized within a database.

The paper develops further the concept of evolving connectionist systems, and one particular model – evolving fuzzy neural networks, that are applied on pattern classification tasks of multimedia data [3]. The evolving systems learn in an on-line, life-long learning mode and adapt to the new data. This mode is crucial when the system is required to adapt quickly to new data and be able to generalize immediately afterwards. The paper demonstrates that this methodology can be successfully applied to on-line pattern recognition from fast streams of multimedia data. It produces high classification accuracy, learns quickly and does not suffer from catastrophic forgetting.

Dilek Kucuk, N. Burcu Ozgur, Adnan Yazici, Murat Koyuncu, present a fuzzy conceptual data model for multimedia data which is also generic in the sense that it can be adapted to all multimedia domains. The model takes an object-oriented approach and it handles fuzziness at different representation levels where fuzziness is inherent in multimedia applications and should be properly modeled. The proposed model also has the nice feature of representing the structural hierarchy of multimedia data as well as the spatial and temporal relations of the data. The model is applied to the news video domain and implemented as a fuzzy multimedia database system where it turns out to be effective in representing this domain and thereby provides an evidence for the general applicability of the model [4]. In this paper, we present a fuzzy conceptual data model for multimedia data and its application to news video domain. The proposed model is also generic in the sense that it could easily be adapted to any multimedia domain. It takes an object oriented approach with the ability to handle fuzziness at the attribute, object/class and class/superclass levels.

In this paper proposed a new rules generation for multimedia data classifying in collaborative environment. ROSETTA tool is applied to verify the reliability of the generated results. The experiments show that the rough sets theory based for multimedia data classifying is suitable to be executed in web services environment. In this paper, proposed a new model for clustering multimedia data based on rough sets theory. The model attempts to classify the multimedia object into three standard multimedia data types. The experimental results have demonstrated that the rough set theory is effective to classify the multimedia data into their respective clusters. The accuracy level of classification result reached 98%. The classification process and results are validated with ROSETTA software. The proposed model has been implemented under web services platform, where J2EE is employed to develop the application. The developed application promises independencies of communication between multi-platform environments. Currently, a comprehensive experiment is under way to combine the concept of temporal based data management to the proposed model in order to improve the classification efficiency [5].

III. METHODOLOGY

In this section we propose a effective methodology to extract the metadata from web multimedia files and classify them based on the extracted metadata by applying data mining techniques. For experimental purpose, out of the total metadata dataset, 60% are used for training and remaining 40% are used for testing the classification model built using Decision Tree and KNN classification methods. The results are analyzed and the efficiency of the proposed method has been demonstrated. The system model of the proposed system is represented in Figure 2. It consists of the following components:

i) Web multimedia-video metadata extraction
ii) Classification model
iii) Classification analysis

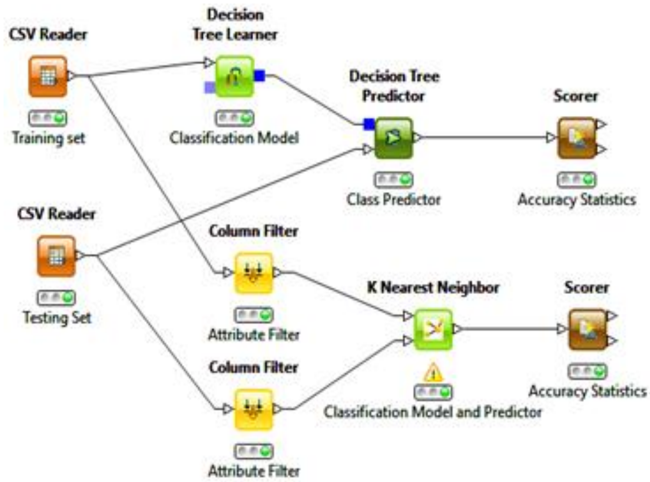

Figure 1: System model of the proposed methodology

The functionality of each component of the proposed system model is discussed in the following subsections.

3.1 Web Multimedia-Video Metadata Extraction and Pre-processing

The metadata of web multimedia-video data are extracted using Mediainfo Extractor tool. Through experimental observation out of 21 attributes 16 attributes found significant for the proposed work. The metadata attributes such as codec id/info, frame rate mode, color space, scan type and compression mode will be excluded during the experiment because the values of these metadata are constant for each tuple. The remaining sixteen metadata - video duration, video bit rate kbps, maximum bit rate kbps, width pixels, height pixels, display aspect ratio, bits/(pixel*frame), stream size mib, audio duration, audio bit rate kbps, maximum bit rate kbps, stream size mib, image resolution, image height, image width, class. The extracted metadata will be store in the form of CSV data file for experimental purpose. The data are pre-processed for filling missing data with mean or mode of each attribute. The Algorithm 1 is describes for the web multimedia-video metadata extraction process.

Algorithm: Web_Multimedia_Metadata_Extraction(WV1, WV2, WV3……WVn)
Input: Multimedia URL
Output: Metadata of Multimedia files
Algorithm
For each selected Multimedia files
Give the selected Multimedia file as input to MediaInfo Extractor
Extract metadata in text format
Store metadata in database

Algorithm 1: Web Multimedia metadata extraction module

3.2 Classification Model

In this experiment we adopt two classification model to classify web multimedia video data. The classification accuracy and efficiency will depend on the constructed classification model. This section represents detailed procedure to construct DT and KNN classification model.

3.2.1 Decision Tree Classification Model

The Decision Tree classification model consist of two major steps i) Attribute selection measures ii) Classification rules. The efficiency of the classification result largely depends on the classification model itself. Hence, construction of robust classification model plays important role in classification. The classification model construction for web multimedia-videos are discussed in the following subsections.

i) Attribute Selection Measures

The attribute selection measures provide specific criteria for each attribute describing the given tuples. As discussed in section 3.1, sixteen attribute class labels are considered for the dataset selected, and are listed in Table 1.

Table 1: Attribute selection for classification

Sl.No	Multimedia metadata Attribute	Descriptions
1	Video Duration	Duration of Video component in times
2	Video Bit rate kbps	Bit rate of video component in kbps
3	Maximum bit rate kbps	Maximum bit rate of video component in kbps
4	Width Pixels	Width of video component in pixels
5	Height Pixels	Height of video component in pixels
6	Display aspect ratio	Display aspect ratio of video component
7	Bits/(Pixel*Frame)	Bits (quality) of video component in pixel
8	Stream size MiB	Stream size of video component in Mebibyte
9	Audio Duration	Duration of Audio component in times
10	Audio Bit rate kbps	Bit rate of audio component in kbps
11	Maximum bit rate kbps	Maximum bit rate of audio components in kbps
12	Stream size MiB	Stream size of audio component in Mebibyte
13	Image Resolution	Resolution of image component
14	Image Height	Height of image component in pixels
15	Image Width	Width of image component in pixels
16	Class	Three different multimedia (Video Domain Specific) classes 1. Entrainment 2. News 3. Sports

The procedure to measure attribute selections for the web multimedia metadata are discussed as follows: The training set D, of class-labeled tuples randomly selected form web multimedia metadata database. The class label attribute has three distinct values namely, 'News', 'Entrainment', and 'Sports'; therefore, there are three distinct classes (i.e., m=3).

Let class C1 correspond to News, class C2 correspond to Entrainment and class C3 correspond to Sports. There are 100 tuples of class news, 62 tuples of class entrainment and 85 tuples of class sports. A node N is created for the tuples in D. To find the splitting criterion for these tuples, the information gain of each attribute is computed as follows.

$$\text{Info (D)} = -\sum_{i=1}^{M} \frac{CiD}{D} \log_2\left(\frac{CiD}{D}\right) \ldots\ldots\ldots \text{(1)}$$

$$\text{Info (D)} = -\frac{100}{247} \log_2\left(\frac{100}{247}\right) - \frac{62}{247} \log_2\left(\frac{62}{247}\right) - \frac{85}{247} \log_2\left(\frac{85}{247}\right) = 0.666$$

Next, to compute the expected information requirement for each attribute. For the attribute *image resolution (IR)* to look at the image resolution size for each category of *IR*. For the *Image resolution* of news category there are out of 43 metadata tuples 35 tuples belongs the IR size 450x360. The IR of *entrainment* category there are out of 38 metadata tuples 38 tuples belongs the IR size 640x360. In the *sports* category there are out of 48 metadata tuples 26 tuples belongs the IR size 1182x720. Using Equation the expected information needed to classify a tuple in *D* if the tuples are partitioned according to *image resolution* is:

$$\text{Info\_A(D)} = \sum_{i=1}^{n} \frac{\text{ID}_i\text{I}}{\text{IDI}} * \text{Info}(\text{D}_i) \ldots\ldots\ldots\ldots\ldots\ldots\text{(2)}$$

$$\text{Info}_{IR}\text{ (D)} = -\frac{100}{247}\left(-\frac{35}{43}\log_2\left(\frac{35}{43}\right) - \frac{8}{43}\log_2\frac{8}{43}\right) + -\frac{62}{247}\left(-\frac{38}{38}\log_2\left(\frac{38}{38}\right) - \frac{0}{38}\log_2\frac{0}{38}\right)$$

$$+ -\frac{85}{247}\left(-\frac{26}{48}\log_2\left(\frac{26}{48}\right) - \frac{22}{48}\log_2\frac{22}{4483}\right)$$

$$= 0.078+0.179+0.240$$

$$= 0.497$$

Hence, the gain in information from such a partitioning would be

$$\text{Gain (A)} = \text{Info (D)} - \text{Info\_A(D)} \quad \ldots\ldots\ldots\text{(3)}$$

$$Gain(IR) = Info(D)\text{-}Info_{IR}(D) = 0.666\text{-}0.497 = 0.169 \text{ bits}$$

Similarly, compute *Gain* (*all attributes*) and get the highest information gain among the attributes, which is selected as the splitting attribute. Node *N* is labeled with *IR*, and branches are grown for each of the attribute's values as shown in table 2.

Table 2: Information Gain of Multimedia Metadata Attributes

Sl.No	**Multimedia metadata Attribute**	**Information Gain**
1	Video Duration	0.123
2	Video Bit rate kbps	0.134
3	Maximum bit rate kbps	0.138
4	Width Pixels	0.113
5	Height Pixels	0.121
6	Display aspect ratio	0.152
7	Bits/(Pixel*Frame)	0.141
8	Stream size MiB	0.148
9	Audio Duration	0.123
10	Audio Bit rate kbps	0.130
11	Maximum bit rate kbps	0.147
12	Stream size MiB	0.149
13	Image Resolution	**0.169**
14	Image Height	0.103
15	Image Width	0.115

The tuples are then partitioned accordingly, where, D_i contains 16 attributes which are outcomes of data partitions D_1, D_2, D_3…Dn, and Info (D_i) can be calculated by using Eq (1). Using Eq(1),(2) and (3) information gain of each attribute will be calculated and the attribute which has highest information gain will be labeled as splitting node[7]. The Table 1 represents the gain obtained by the Decision Tree classification model, in which the attribute 'Image Resolution' has the highest gain among the selected attributes. Hence, the attribute 'Image Resolution' is selected as root node of the tree. In the similar way, at each point of node, the gain will be calculated and tree will be formulated accordingly.

ii) Classification Rules

Classification rules can be extracted from the tree structure of the classification model for the dataset chosen as shown in Figure 2.

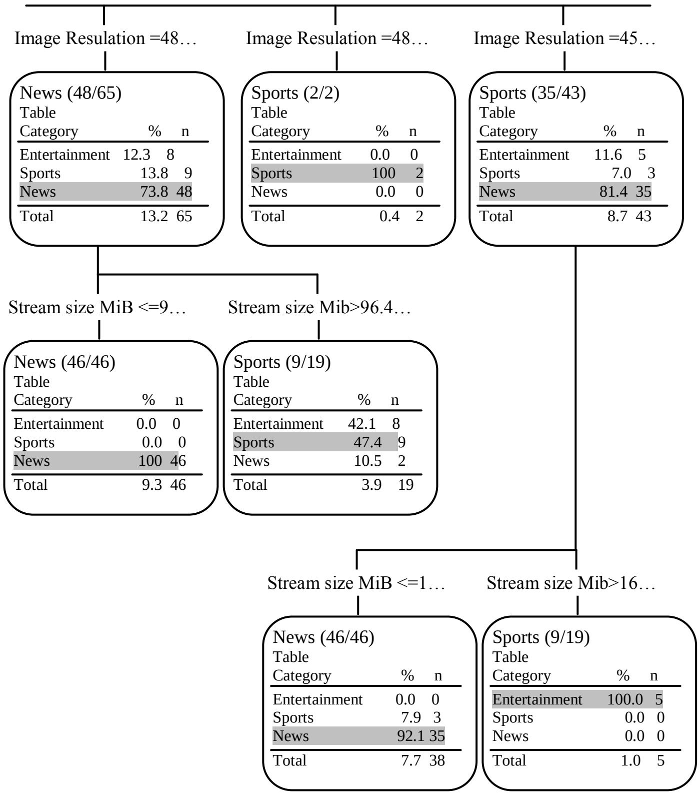

Figure 2: Tree structure result of DT classification model

The above tree can be converted to classification rules by traversing the path from root node to each leaf node in the tree. In Figure 2, the root node is created with the splitting values of the attribute 'Image Resolution'. Each node contains the information of class label in terms of correctly classified instances and incorrectly classified instances. The classification rules extracted from class predictor as shown in Figure 3.

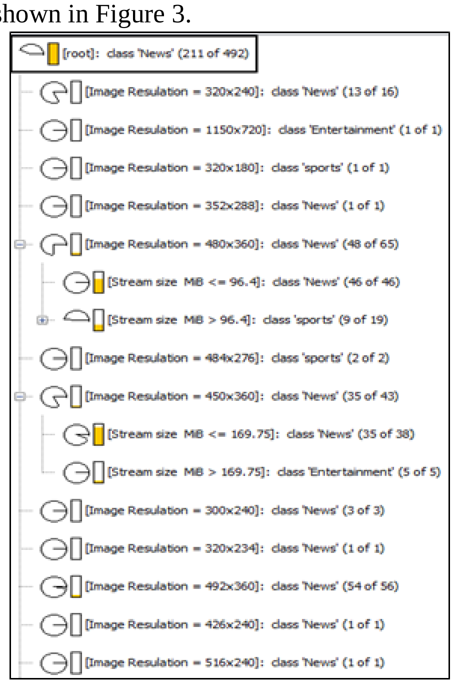

Figure 3: Classification rules

3.2.2 K-Nearest Neighbor Classification Model

The k-Nearest Neighbor Classification model is based on learning by analogy, that is, by comparing a given test example with training examples that are similar to it. The training examples are described by 16 attributes. Each example represents a point in a 16-dimensional space. In this way, all of the training examples are stored in a 16-dimensional pattern space. When given an unknown example, a k-nearest neighbor algorithm searches the pattern space for the k training examples that are closest to the unknown example. These k training examples are the k "nearest neighbors" of the unknown example. "Closeness" is defined in terms of a distance metric, such as the Euclidean distances are shown below.

$$dist(X_1, X_2) = \sqrt{\sum_{i=1}^{n} (x_{1i} - x_{2i})^2} \quad \ldots\ldots(4)$$

In this experiment, For nominal attributes News, Entertainment, Sports a method is to compare the corresponding value of the attribute in tuple X_1 with that in tuple X_2. If the two are identical (tupls X_1 and X_2 both have the same class), then the difference between the two is taken as 0. If the two are different (tupls X_1 is news but tuple X_2 is entertainment), then the difference is considered to be 1. The set of test data based on the K-Nearest Neighbor that, out of 247 instances, 231 tuples are correctly classified and 16 tuples are incorrectly classified by the KNN classification model.

3.3 Classification Analysis

In this section, performance evaluation measures such as TP, FP, precision, recall and F-Measure will be calculated to measure classification accuracy and efficiency of DT and KNN classification model. Also the classification accuracy of DT and KNN will be compared. The quality of the DT and KNN classification models will be represented in the form of confusion matrix.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

4.1 Classification using Decision Tree model

To test the efficiency of the classification models constructed using Decision tree and KNN, the multimedia dataset is extracted from the data mining tool which consists of 247 web multimedia- video metadata instances. The performance of the model is measured in terms of number of correctly classified instances, number of incorrectly classified instances, TP rate, FP rate, precision, recall and F-score. The Table 3 represents classification result obtained by the Decision Tree classification model.

Table 3: Classification result of Decision Tree classification model

Sl.No	*Class Labels*	*Total Weight*	*Correctly Classified*	*Incorrectly Classified*	*TP*	*FP*	*Precision*	*Recall*	*F-Measure*
1	News	100	92	8	92	5	0.948	0.92	0.934
2	Entertainment	62	59	3	59	11	0.843	0.952	0.894
3	Sports	85	77	8	77	3	0.962	0.906	0.933
	Total	***247***	***228***	***19`***	***228***	***19***	***0.917***	***0.926***	***0.921***

It is observed from the Decision tree experimental result that, out of 247 instances, 228 tuples are correctly classified and 19 tuples are incorrectly classified by the Decision tree classification model. The class labels 'Sports' has highest precision and accuracy. Also the falls positive rate of 'sports' is very less with respectively. In the 'Entrainment' Class label out of 62 records 59 are correctly classified and 3 were incorrectly classified by DT model. However the falls positive rate of class label 'Entrainment' is high as compare to remaining class label. The overall efficiency of Decision tree classification is found 92.1%.

4.2 Classification using KNN model

The K-Nearest Neighbor experimental result that, out of 247 instances, 231 tuples are correctly classified and 16 tuples are incorrectly classified by the KNN classification model. The class labels 'Sports' has highest precision and accuracy. Also the falls positive rate of 'sports' is very less with respectively. In the 'News' Class label out of 100 records 97 are correctly classified and 3 were incorrectly classified by KNN model. However the falls positive rate of class labels 'News' is high as compare to remaining class label. The overall efficiency of KNN classification is found 93.1%. The Table 4 represents classification result obtained by the Decision Tree classification model.

Table 4: Classification result of KNN classification model

Sl.No	*Class Labels*	*Total Weight*	*Correctly Classified*	*Incorrectly Classified*	*TP*	*FP*	*Precision*	*Recall*	*F-Measure*
1	News	100	97	3	97	5	0.951	0.97	0.96
2	Entertainment	62	57	5	57	7	0.891	0.919	0.905
3	Sports	85	77	8	77	4	0.951	0.906	0.928
	Total	***247***	***231***	***16***	***231***	***16***	***0.931***	***0.931***	***0.931***

The quality of the DT and KNN model is represented the terms of confusion matrix and which is represented in table 5 as a conclusion of analysis of classification result obtained by the DT and KNN model. The table 5 represents comparison of efficiency from the comparison analysis of classification
model is found better to classify web multimedia-video data.

Table 5: Confusion matrix of DT and KNN classification result

Confusion Matrix							
Decision Tree Classification Model				K-Nearest Neighbor Classification Model			
===Confusion Matrix===				===Confusion Matrix===			
a	**b**	**c**	**←Classified as**	**a**	**b**	**c**	**←Classified as**
92	5	3	| a=News	97	3	0	| a= News
3	59	0	| b= Entrainment	1	57	4	| b= Entrainment
2	6	77	| c=Sports	4	4	77	| c= Sports

The Graphical representation of comparison analysis is represented in Figure-3.

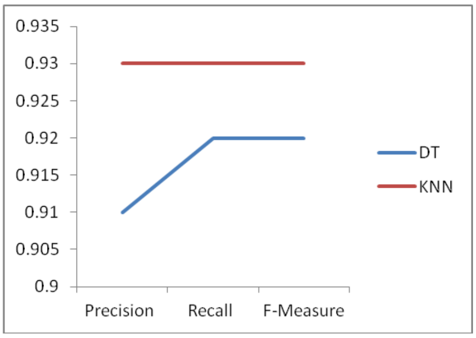

Figure 2: Comparison of classification result

The experimental result shows that, KNN classification model works well as compared to DN classification model. The web multimedia-video metadata datasets contains all independent attribute as continuous values. Due to this factor the DT classification has less accuracy than KNN classification model.

V. CONCLUSION AND FUTURE WORK

In this work, we classified web multimedia-videos based on their category using web multimedia metadata. The web multimedia-video metadata are extracted and stored in a database for classification. The Decision Tree (DT) and KNN classification algorithms are chosen to classify the web multimedia-videos. The classification results of DT and KNN classification models are compared and found KNN classification model is more efficient for classify web multimedia-videos using metadata. Also the Decision tree classification model has less efficiency web multimedia-video based on independent attributes. The accuracy of Decision Tree classification model if found 92.1% and the accuracy of KNN classification model is found 93.1%. The future work is to improve the classification accuracy of DT classification model on the classification of web multimedia-videos.

REFERENCES

[1] Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, "The New Data and New Challenges in Multimedia Research", arXiv:1503.01817v1 [cs.MM] 5 Mar 2015.

[2] D. M. SHOTTON*, A. RODRÍGUEZ†, N. GUIL† & O. TRELLES†, "A metadata classification schema for semantic content analysis of videos", Journal of Microscopy, Vol. 205, Pt 1 January 2002, pp. 33–42

[3] 1NIKOLA KASABOV, 1IRENA KOPRINSKA, 2GEORGI ILIEV, "Evolving Connectionist Systems for On-Line Pattern Classification of Multimedia Data", 9th WSEAS International Conference on NEURAL NETWORKS (NN'08), Sofia, Bulgaria, May 2-4, 2008, ISBN: 978-960-6766-56-5.

[4] Dilek Kucuk, N. Burcu Ozgur, Adnan Yazici, Murat Koyuncu, "A Fuzzy Conceptual Model for Multimedia Data with Application to News Video Domain", L. Magdalena, M. Ojeda-Aciego, J.L. Verdegay (eds): Proceedings of IPMU'08, pp. 1741-1748, Torremolinos (M_alaga), June 22-27, 2008.

[5] M. Nordin A. Rahman, Yuzarimi M. Lazim, Farham Mohamed, Syarilla Iryani A. Saany and M. Kamir M. Yusof, " Rules Generation for Multimedia Data Classifying using Rough Sets Theory", International Journal of Hybrid Information Technology, Vol.6, No.5 (2013), pp.209-218.

[6] ChenGang "MediaInfo extractor – A Tool for Media Data Mining", 2011. http://mediaarea.net/en/MediaInfo.

[7] Siddu P. Algur, Prashant Bhat, Suraj Jain, "The Role of Metadata in Web Video Mining: Issues and Perspectives", International Journal of Engineering Sciences & Research Technology, February-2015.