

RT AND J48 CLASSIFICATION MODELS FOR PREDICTING ABNORMAL WEB MULTIMEDIA DATA

¹SIDDU P. ALGUR, ²SURAJ JAIN, ³PRASHANT BHAT

^{1,2} Department of Computer Science, Rani Channamma University, Belagavi-591156, Karnataka, India

³ School of Computational Sciences and IT, Garden City University, Bengaluru

E-mail: ¹siddu_p_algur@hotmail.com, ²suraj_point@gmail.com, ³prashantrcu@gmail.com

Abstract - Now a days, the 'Data Science Engineering' becoming emerging trend to discover knowledge from web multimedia such as- YouTube multimedia, Yahoo Screen, Face Book multimedia etc. Petabytes of web multimedia are being shared on social websites and are being used by the trillions of users all over the world. Recently, discovering outliers among large scale web multimedia have attracted attention of many web multimedia mining researchers. There are plenty of outliers/abnormal multimedia exists in different category of web multimedia. The task of classifying and prediction of web multimedia as- normal and abnormal have gained vital research aspect in the area of Web Mining Research. Hence, we propose novel techniques to predict outliers from the web multimedia dataset based on their metadata objects using data mining Decision Tree algorithms such as Random Tree (RT) and J48 Tree algorithms. The results of Random Tree and J48 Tree classification models are analyzed and compared as a strategy in the process of knowledge discovery from web multimedia.

Keywords - Outliers, Decision Tree, J48 Tree, Web Multimedia Outliers, Prediction, Knowledge Discovery.

I. INTRODUCTION

YouTube is recognized as one of the most successful user-generated multimedia sharing sites nowadays. YouTube has over a billion users — almost one-third of all people on the Internet — and every day people watch hundreds of millions of hours on YouTube and generate billions of views [1] [2]. In order to facilitate users to find interesting multimedia from a large number of multimedia, YouTube provides different features/metadata objects such as – view counts, rate, ratings, number of comments, favorites, key words, information regarding likes and dislikes etc.

The objective of this study is to classify and predict outlier (abnormal) multimedia among large scale web multimedia using their metadata objects. To thrive in the proposed objective of the work, large scale web multimedia metadata objects are extracted from the standard YouTube dataset website [3]. This metadata objects includes various attributes such as- 'Category', 'Length', 'Views', 'Rate', 'Avg Ratings' 'Number of Comments' and 'Outlier (Nominal)' of each web multimedia.

The main contributions of our work are as follows:

- Web Multimedia Metadata Object dataset extraction and effective preprocessing for the experiment.
- The analysis and knowledge discovery process from the results of 10 cross validation classification and predictions of outliers by the built Random Tree (RT) and J48 Tree models.

Many outlier models/algorithms and data mining machine learning tools are developed in recent years. Using different data mining algorithms and machine learning tools such as R programming and WEKA, it

is possible to classify and predict outliers from the web multimedia based on their features/metadata objects.

The rest of the paper is organized as follows: The section 2 represents related works on the clustering of web multimedia, section 3 represents proposed web multimedia clustering methodology, section 4 represents performance evaluation analysis of outlier models and comparison of efficiency of outlier models, and finally section 5 represents conclusion and future enhancements.

II. RELATED WORKS

This section represents some related previous works which are implemented to find abnormal web multimedia/ abnormal web multimedia events using metadata objects.

The authors Chueh-Wei Chang, et al. [2], proposed a framework for spatial relationship construction, abnormal event detection and multimedia content searching with respect to visual surveillance applications. The proposed system [2] can automatically detect the abnormal events from monitoring areas, and select the representative key frame(s) from the multimedia clips as an index, then store the color features of the suspect objects into the surveillance database. A graph model has been defined to coordinate the tracking of objects between multiple views. This was helpful to the surveillance system to check the route of objects whether go into a critical path or not. A variety of spatio-temporal query functions can be provided by using this spatial graph model. To achieve the content-based multimedia object searching, a kernel-based approach has been employed as a similarity appraise

