Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Video Content-Based Advertisement Recommendation System using Classification Technique of Machine Learning

To cite this article: R C Konapure and L M R J Lobo 2021 *J. Phys.: Conf. Ser.* **1854** 012025

Video Content-Based Advertisement Recommendation System using Classification Technique of Machine Learning

R C Konapure[1] and Dr L M R J Lobo[2]

[1]PG Student, Department of Computer Science and Engineering, Walchand Institute of Technology, P.A.H. Solapur University, Solapur, Maharashtra, India

[2]Professor, Department of Computer Science and Engineering, Walchand Institute of Technology, P.A.H. Solapur University, Solapur, Maharashtra, India

[1]konapurer@gmail.com, [2]headitwit@gmail.com

Abstract Content-based advertising is a method by which we advertise on a video media based on a relevant topic assigned to the video. In digital advertising, the advertisements shown to a user is based on the user's behaviour on the internet. Streaming platforms are then used to target audience based on parameters like user's geo-location, interests, watch history, age, etc. In most cases, the advertisements shown are not relevant; an undesired impact is created. Content-based advertising helps to convey the message with increased efficiency and simultaneously optimizes its conversion rate. In this proposed system we take the video metadata as input and apply the NLP techniques for text classification which categories the video and assigns a relevant advertisement to it. The second module takes the video as an input. Thereafter the video is converted into N individual frames to tackle the video classification as an image classification problem. In this proposed system we train a Convolutional Neural Network to identify the topic of the video on an image dataset and compare its performance with a pre-trained model. We create the image dataset by downloading images from the internet. We also create a video advertisement's dataset by web scrapping. This proposed system makes sure that the user is shown the advertisement in reference to the video. This increases probability of the user visiting the client's website.

1. Introduction

Present day digital-age advertising is more complex than ever. Advertisers today have a chance to meet consumers at every stage of their online browsing process. However, it takes a deep level of insight into conversion behaviour and easy access to real-time digital decision-making tools to know when, how and to what extent it is indeed needed to communicate with a customer. Today's modern advertisers must rely on technology to decide the best way to boost their advertising return on investment (ROI) and meet their target clients. Recommendation systems have been implemented with the emergence of streaming sites like YouTube, Netflix and many other such web services. Recommendation systems are inevitable in our everyday online browsing, from e-commerce (to recommend a similar product) to online ads (based on a user interests).

1.1 Motivation

The social advertising market is ever growing, having an outreach to customers of all the age groups and remote areas. The reason behind the use of social networking sites as a platform for advertising is to suggest a potential customer who is interested in a product to get relevant information resulting in a high conversion ratio. An advertiser is motivated to use a social networking site as standardised platform. The 2018 State of Digital Advertising Study from Adobe Digital Insights reveals that social media advertisements had drawn three times more non-consumers to the retailer's website by the end of 2017 than current customers. Today's people concentrate more on online facilities for advertisement which make their life easier. As this survey report brings to light that people connect to these sites to a greater extent which further helps the advertiser. The video accounts for 80% of web traffic over the course of 2020, according to the expert's estimate. YouTube is well known publically in the digital market involving video presentations. The advertisements occupy less memory space and are optimised. The user base is almost as high as that of Facebook, with approximately over 2 billion monthly active users. The difference between the YouTube and Facebook models is that for extended sessions, every single person watches videos. This is exactly what is expected. There was a gain for motivation due to the following:

1. The proposed system helps to provide an accurate advertisement during the video which consequently increases the actual conversion ratio of the targeted audience.

2. As we proposed to provide the advertisement in reference to the video content, users will not be displeased by irrelevant advertisements.

1.2 *Objective*

Paid digital advertisement has shown great growth continually on various platforms. Most marketers realize the importance of fulfilling a customer's requirement, thus they use these platforms to reach the core audience. The methodologies adapted in the algorithms are based on video content. This automatically decreases the load on a server. A reflection is seen on the sorted data which reduces unnecessary searches, thus saving access time. Thrust technologies like machine learning and pattern matching are adopted to generate correct customer through the metadata of advertisements in the videos. This ensures that an advertisement will be shown only to relevant users who will benefit from the system.

2. Literature Review

To understand the problem the reviews of the previous studies in this domain need to be visited. The methodology followed to identify the unexplored part of the field of study under consideration is based on extensions provided in this literature.

Onur Sevli et.al. [2] focused his study on recommendation systems w.r.t advertisements which suggested the use of a Twitter platform to show the advertisement to the correct user. For this task, post shares, news by a user community in Turkish, natural language processing and big data analysis techniques were used to distribute ads only to those users who need the product. This developed system described the fields of personal interest by identifying the word patterns most widely used in sharing. A web service was developed to present content to consumers, marked by a range of advertisement database categories and keywords and ideally tailored to user preferences. Abu Bashar et al. [3] explored the experience of consumers and advertisers in the state of Punjab on a social networking platform. This study resulted in the suggestion that the metrics for the efficacy of social media should be interesting, insightful, interactive and accurate. Social marketing sites need to be alert to consumers' shifting tastes and expectations. This study indicated that advertising encourages market competitiveness that increases the availability of customers with higher quality goods. Thirumalaisamy Ragunathana et al. [9] suggested a model of customer behaviour based on posting a relevant advertisement after a customer's visit to a website. The proposed model analyses the mechanism and activities in which individuals browse, pick, purchase, use, review and dispose products and services in order to fulfil their needs and desires. The author used Hadoop to handle huge data of the customer behavioural model used to display relevant advertisements on the website.

A multi-label video identification system was proposed by Kwangsoo et. al. [10] on the YouTube data. The growth of video data is rapid because of developments in digital technology. Therefore, the need for techniques to automatically classify visual content is increasing. The YouTube-8M dataset along with NetVLAD and NetFV models as well as the Huber loss function was used to check the needs for video classification problems. Mariana Arantes et. al. [4] addressed the use of YouTube video advertising. YouTube is the most popular streaming site that uses advertisements to raise billions of dollars each year. On YouTube, individuals can create and upload their content after they hit a certain number of views and subscribers the channel can be monetized. This research provided insights into how ads generate revenue, the popularity of video ads, user behaviour when they are exposed to advertisements, and how content creators can generate revenue as a person via their YouTube channel. This research focused primarily on user behaviour, with an emphasis on video content. Dardis et al. [11] experimented in order to understand the influence of banner advertising and video ads on brand recall. Within two distinct game environments, they conducted the study: games developed for advertising purposes (called advergames) and non-branded games. Among their results, they found that in non-branded games, video ads are better than banner ads and also that the role of mid-roll video ads was more prominent. Thus, looking into the work done by previous researchers it was concluded that each researcher focused on a very particular video advertisement scenario. This developed motivation to study video advertisements with a broader view.

3. Data Collection

Data collection is an important step to be followed in order to apply machine learning algorithms. Relevant answers to specific questions in the data collection process can determine results from an established system. These are inherently based on collecting and estimating information on targeted variables. A collection of data for experimentation was extracted from scratch and a unique database is created.

3.1 *Text Data*

To perform the text classification, YouTube videos metadata using YouTube API v3 is collected. This dataset consists of fields: video id, title, description, hashtags, duration and upload details etc. A total of more than 10,000 video metadata is collected and divided into six categories. This dataset items contain a lot of noise which proves to be challenging. Before building a classification model we should perform data pre-processing. Natural Language Toolkit (NLTK) package of NLP does the job of pre-processing efficiently. It consists of data pre-processing tools required for text classification.

3.2 Image Data

For the video classification model a Convolutional Neural Network is trained to identify the sport in the video. Image data from Google images of 22 categories of sports (badminton, baseball, cricket, football, etc.) is collected. Each sport category has around 700 – 800 images. The same dataset is used to create a video classification model with Keras and deep learning. A comparison was done between the developed model and the inception v3 model.

3.3 Advertisement data

The main objective of the proposed system is to recommend a video advertisement to the video on a streaming platform based on its content. In order to achieve this various advertisements in the video format are collected and metadata of these are extracted and stored in a database. The advertisement video data consists of information regarding video URL, video id, title, description etc.

4. Methodology

The use of the advertisements depended on how relevant and suitable they would be to a specific user. In this paper, the content match scenario is used and two classification models are proposed. One model for text data and the other for video data.

4.1 The Proposed Recommender Algorithm

Video content pre-dominates the working of the recommender algorithm. Keeping in mind relevance to a user based on the content in a video is of prime importance. This reflects on a user giving a better response to the system. The proposed algorithm solves the problem of recommending more related advertisements and provides an opportunity to generate a higher user response and satisfaction thus resulting in increasing business value [5]. The steps in the algorithms are as follows:

4.1.1 Text classification. The text classification model of the proposed system is developed as shown in Figure 1.

Figure 1: An overview of the text classification

The first step in data pre-processing is to handle the missing data. Since missing values are supposed to be text data in the proposed system there is no way to substitute them, the only option is to remove them. Following this on the remaining data natural language processing text cleaning techniques are performed to generate clean and required data [5]. This approach involves converting to lowercase, removing numerical values and punctuation, removing extra white spaces, tokenizing into words, lemmatization, removing non-alphabetical words and stop words.

Data extraction and data analysis have to be performed. Data cleaning, transformation, and data modelling comprise analysis to discover useful information and to support decision-making. The data extraction process consists of retrieval of information from the available data sources for further data processing [5].

Data modelling is used to define and analyse the data needed to support the machine learning model and store it into a database to perform required operations on it. In data modelling, the processed dataset is divided into a training dataset and a test dataset. In the proposed system we form a training dataset with the features extracted using the Term Frequency Inverse Document Frequency (TFIDF) vectorization method [5].

Classification label is a process of producing labels i.e. output from the classifier model. The labels are related to the content.

4.1.2. Video Classification. The video classification model of the proposed system is developed as shown in Figure 2.

Journal of Physics: Conference Series **1854** (2021) 012025 doi:10.1088/1742-6596/1854/1/012025

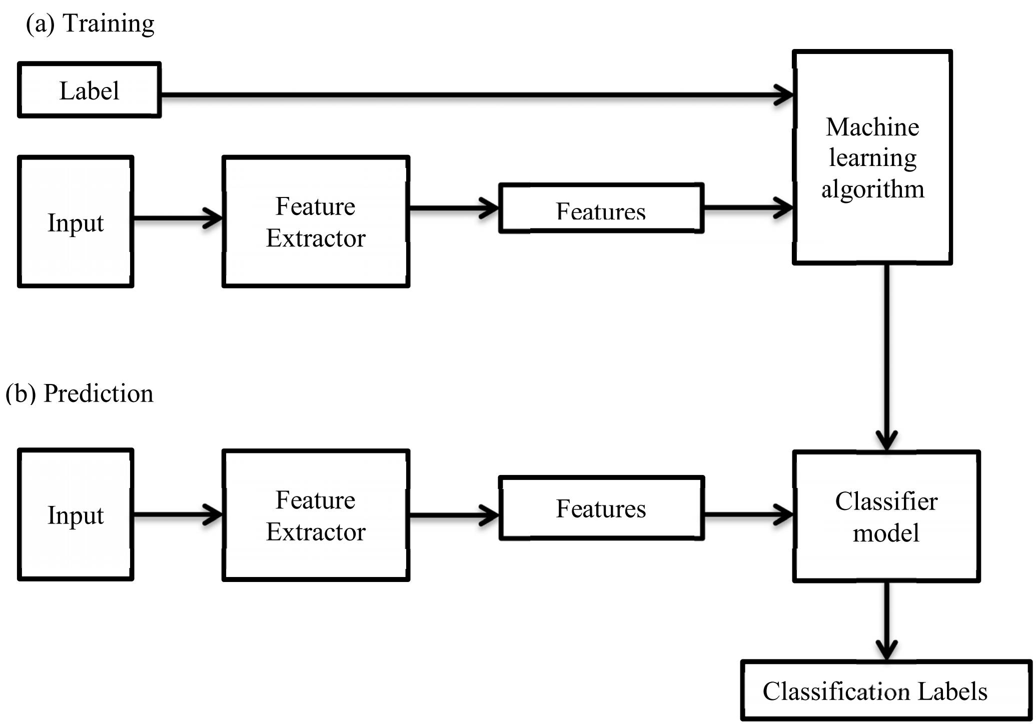

Figure 2: An overview of the video classification

The model is trained on the training dataset using a supervised learning method. The training dataset consists of an input and the corresponding output, which is commonly denoted as the target variable. The current model runs with the training dataset and produces a result, which is then compared with the target variable, for each input in the training dataset. Based on the result of the comparison and the specific learning algorithm being used the parameters of the model are adjusted.

In the prediction stage, a test dataset is taken; this dataset is independent of the training dataset. The prediction phase works similarly to the training phase with the difference that in the training set pre-defined labels to the machine learning classification algorithm are fed and a classifier model is built. In prediction extracted features to the classifier model are fed which in turn produces labels for classification. The test dataset is a set of examples used to check the performance of a producing classifier model and results are achieved in terms of labels for classification.

To summarize the whole process videos are first converted into a series of images which are passed through CNN. The prediction from the CNN is obtained and a list for all the frames is determined. The average of these frames is calculated and the labels are chosen having highest probability. Frames are then labeled and outputs are written to the disk.

In order to develop a simple transfer learning with an Inception v3 architecture model, the same data is used. On ImageNet images, the Inception v3 architecture model is trained; a new top layer that can identify other image classes is trained. For each image, the top layer receives a 2048-dimensional vector as the input. On top of this representation, a softmax layer is applied. If the softmax layer includes N labels, this refers to the parameters of the learning N + 2048*N model corresponding to the biases and weights learned.

5. Results and Discussion

5.1 Text Classification

The testing starts with the data of six categories Travel Blogs, Science and Technology, Food, Manufacturing, History, Art, and Music. The classifier used in the experiment is the LSTM classifier.

Table 1: Confusion Matrix

	art and music	food	History	manufact uring	science and technology	travel
art and music	394	2	3	3	0	1
food	5	425	0	1	6	1
history	6	0	411	0	2	2
manufacturing	2	0	0	395	0	0
science and technology	5	1	2	4	391	0
travel	7	6	0	0	0	425

The confusion matrix is built by LSTM Classifier as shown in Table 1. Looking vertically at the value of a class in the confusion matrix, one can see the instances of a category as assigned by a classifier. For example, looking at the above confusion matrix shows that the LSTM classifier has correctly classified 394 instances of art and music as the Health class (True Positives, TP) while 25 instances are False Positives (FP).

Table 2: text classification accuracy by class

	precision	recall	f1 - score	support
art and music	0.94	0.98	0.96	403
food	0.98	0.97	0.97	438
history	0.99	0.98	0.98	421
manufacturing	0.98	0.99	0.99	397
science and technology	0.98	0.97	0.98	403
travel	0.99	0.97	0.97	438
accuracy			0.98	2500
macro avg	0.98	0.98	0.98	2500
weighted avg	0.98	0.98	0.98	2500

Considering precision, recall, f1-score, support it can be seen that the LSTM classifier consistently gives better performance. For the six categories, the Advertisement Recommendation system dynamically recommended the relevant advertisements up to 98%.

5.2 Video Classification

Due to the high cost of processing time and processing power required by most of the classifiers, and the number of iterations to produce the best from the classifiers the experiment was conducted in stages.

Figure 3: Output of CNN model

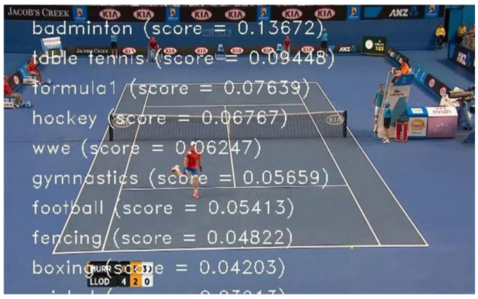

Figure 4: Output of the pre-trained model

Figure 3 shows the output of the video classification model with a classification label. Using this classification label as a keyword we retrieve the advertisement's related to tennis.

Table 3: video classification accuracy by class

	precision	recall	f1 - score	support
football	0.88	0.95	0.92	196
tennis	0.90	0.91	0.90	179
weight_lifting	0.97	0.84	0.90	143
accuracy			0.91	518
macro avg	0.92	0.90	0.91	518
weighted avg	0.91	0.91	0.91	518

In Table 3 considering precision, recall, f1-score, support it can be seen that the video classifier consistently gives better performance with 91 % accuracy. We compared the above results with the pre-trained model's results. Table 4 below shows the labels with their calculated score of pre-trained model.

Table 4: Pre-trained model results comparison

tennis (score 0.09849)	fencing (score 0.05321)	kabaddi (score - 0.03406)	wrestling (score 0.01801)
table tennis (score 0.09332)	hockey (score 0.04715)	baseball (score - 0.03229)	basketball (score 0.01588)
gymnastics (score 0.09309)	chess (score 0.04510)	shooting (score 0.03163)	ice hockey (score 0.01523)
wwe (score 0.09008)	football (score 0.04218)	volleyball (score 0.02345)	motogp (score. 0.01435)
formula1 (score 0.07433)	swimming (score 0.03859)	weight lifting (score 0.02197)	
boxing (score 0.06044)	cricket (score - 0.03547)	badminton (score- 0.02169)	

5.3 Recommendation System Results

Figure 5 shows the advertisement generated from the keyword identified by CNN model. The keyword is then retrieved from the advertisement dataset by matching title, description of the source video. Only matched data will return more relevant videos. In the proposed system the CNN model identifies the sport as tennis. Using this keyword this system retrieves the video shown is Figure 5.

Figure 5 : Advertisement shown related to tennis

6. Conclusion

In the proposed system with the text classification model advertisement placement on a video has achieved an accuracy of 98% with LSTM classifier which was better than Naïve Bayes which gave an accuracy of 96% and Adaboost Classifier which gave an accuracy of 86% for the same input data. In video classification the results of the developed system have been compared to the pre-trained model. Keeping the dataset constant the developed model gave an accuracy of 91% which is calculated on average score of the labels for a fixed number of frames. However for pre-trained model the accuracy score has been calculated for individual frames. Though they have reached an accuracy of 98% this does not consider the average scores of labels. This accuracy may fall down after achieving all the labels once all the frames are available for analysis.

With the proposed system, advertisers are ultimately benefited from the outcome of this research effort. They are able to provide their consumers with more effective advertisement campaigns while generating more revenue for both parties. This model uses contextual data to publish advertisements more efficiently increasing user response to advertising through better placement, duration, and format of advertisements to the targeted audience. This system provides accurate and more related advertisements and information to users more purposefully. This analysis can also be used to improve other aspects of the online advertising business facilitating them to gain a competitive advantage in the growing and competitive market. The suite of algorithms that have been applied resulted in successful and acceptable quality predictions.

7. References

[1] Virda S, Yu-Qian Z, Achmad Nizar H, Puspa I and Bo H. *International Journal of Information Management 48* 2019 Exploring the psychological mechanisms from personalized advertisements to urge to buy impulsively on social media. 96–107.

[2] Onur S, Ecir K. *International Journal of Tehnički vjesnik* 2015 Advertising recommendation system based on dynamic data analysis on Turkish speaking twitter users ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) Retrieved from https://hrcak.srce.hr/file/265215.

[3] Abu B, Irshad A. *ELK Asia Pacific Journal of Marketing & Retail Management* Effectiveness of Social media as a marketing tool: An empirical study.

[4] Mariana A, Flavio F and Jussara M. *The Journal of Web Science* Volume 4 2018 Towards Understanding the Consumption of Video-Ads on YouTube. 1–19

[5] Rushikesh K, Dr. L.M.R.J. Lobo. *MAT Journals* 7 Feb. 2020 Text Data Analysis for Advertisement Recommendation System Using Multi-label Classification of Machine Learning.

[6] Sree Vani M. *International Journal of Advanced Research in Computer and Communication Engineering* Vol. 5, Issue 2, February 2016 A Recommender System for Online Advertising. 599-604

[7] Nhan N-T, Dana M, Kinda K, David R, Flavian V, Elena Simona L, Steven M, Dominique Q. *arXiv:1909.04190v1 [cs.IR]* 9 Sep 2019 Recommendation System-based Upper Confidence Bound for Online Advertising.

[8] Missi Hikmatyar and Ruuhwan. *Journal of Physics: Conference Series 1477 032024* 2020 Book Recommendation System Development Using User-Based Collaborative Filtering.

[9] Thirumalaisamy R, Sudheer K, Battulab V. *International Journal of Computer Science Procedia* 50 Advertisement Posting based on Consumer Behavior. 329 – 334

[10] Kwangsoo S, Junhyeong J, Seungbin L, Boyoung L, Minsoo J, Jongho N. *IEEE Conference on Computer Vision and Pattern Recognition* Approach for video classification with multi-label on YouTube-8M dataset 5297-5307.

[11] Dardis, F. E., M. Schmierbach, B. Sherrick, F. Waddell, J. Aviles, S. Kumble, and E. *Journal of Interactive Advertising Bailey* Adver-Where? Comparing the Effectiveness of Banner Ads and Video Ads in Online Video Games. 1–14

[12] Arantes, M., F. Figueiredo, and J. M. Almeida. *WebSci* Understanding video-ad consumption on YouTube: a measurement study on user behaviour, popularity, and content properties.

[13] Li, H., and H.-Y. Lo. *Journal of Advertising* 44(3) Do You Recognize Its Brand? The Effectiveness of Online In-Stream Video Advertisements. 208–218