MOOCS RECOMMENDER BASED ON LEARNING STYLES

Thesis

Project ID: 19-089

Liyanage A.Y.K. | IT 16 0327 98

De Silva W.A.T.P. | IT 16 0519 80

Pathirana P.H.P.S.L. | IT 16 0043 82

Hilmy S.B.M. | IT 16 0053 72

Bachelor of Science Special (Honors) in Information Technology

Specializing in Software Engineering

Department of Software Engineering

Sri Lanka Institute of Information Technology

Sri Lanka

Date of Submission: 2019-05-13

MOOCS RECOMMENDER BASED ON LEARNING STYLES

Project ID: 19-089

Liyanage A.Y.K. | IT 16 0327 98

De Silva W.A.T.P. | IT 16 0519 80

Pathirana P.H.P.S.L. | IT 16 0043 82

Hilmy S.B.M. | IT 16 0053 72

Supervisor: Mr. Nuwan Kodagoda

Co-Supervisor: Ms. Kushnara Suriyawansa

Date of Submission:

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also I hereby grant to Sri Lanka Institute of Information Technology the non-exclusive right to reproduce and distribute my dissertation in whole or part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as article or books).

Name	ID	Signature
Liyanage A.Y.K.	IT16032798	
De Silva W.A.T.P.	IT16051980	
Pathirana P.H.P.S.L.	IT16004382	
S.B.M. Hilmy	IT16005372	

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Signature: Date:

Signature of the supervisor: Date:

Signature of the co-supervisor: Date:

ABSTRACT

Massive Open Online Courses (MOOCs) are a popular learning medium in the current day and age. Its popularity has been increasing steadily over the years and now are helping people all around the world learn at an unprecedented scale. When it comes to learning anything a proven to be effective way of doing this is by discussing the topic with peers. This trait of learning is facilitated by the MOOCs and their platforms in the form of 'Forums'. When it comes to learning through this method or enhancing the learning effectiveness, some people prefer it more than others. In fact, research has shown that as a result of their preference towards learning through discussion they learn faster when engaging in subject content related discourse. Forums in general contain a large amount of information which can be used to produce useful insights to the students of the course. This paper presents the analysis of MOOC forums to generate useful insights which can in turn help students choose the course they are going to follow. Raw forum data is gathered from different MOOC platforms with the use of web crawling techniques. Various analytical methods are used to get useful information from the retrieved data which includes the Natural Language Processing technique (NLP) sentiment analysis.

Keywords: MOOCs, Forum Analysis, Data Mining, Sentiment Analysis

ACKNOWLEDGEMENT

This academic paper and work described in it were conducted as a part of 4th year research module (Comprehensive Design Analysis Project) under the degree program BSc (Hons) in Software Engineering in Sri Lanka Institute of Information Technology. The fruition of this research and the associated software project are results of dedication of the 4 group members and the immense support given by a supervising panel.

Thus, we would like to express our gratitude towards Dr. Nunan Kodagoda and Ms. Kushnara Suriyawansa, who are our supervisors as well as lecturers of Sri Lanka Institute of Information Technology. Completion of this academic work would not have been possible without their guidance and insight. We are extremely grateful to Dr. Malitha Wijesundara and Dr. Dasuni Nawinna, who reviewed and approved our research topic and provided valuable suggestions. Furthermore, we wish to express our gratitude to Mr. Jayantha Amaraarachchi, the Head of SLIIT Centre for Research who provided us with immense knowledge and guidance throughout.

Also, we would like to express our gratitude to all our colleagues and companions who helped us in testing and in various other ways to make our research come to fruition. At the same time, we would like to thank everyone who might not have been mentioned but has given their support and encouragement to us.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVATIONS

MOOC	Massive Open Online Course
SIMD	Single Instruction Multiple Data-stream
ILS	Index of Learning Styles
CNN	Convolutional Neural Network
VM	Virtual Machine
NLP	Natural Language Processing
ML	Machine Learning
VGG	Visual Geometry Group
FSLSM	Felder Silverman Learning Style Model

HCI	Human Computer interaction

1 INTRODUCTION

1.1 Background Literature

The completion rate of MOOCs has been found out to be low by many researches done on the engagement of consumers with MOOCs over the duration of courses [1].Felder-Silverman Learning Style Model [2] has highlighted the importance of acknowledging that an individual can inherit a learning style or few that he or she will be more inclined towards and dimensions that an individual can fall into, such as Active or Reflective, Visual or Verbal, Sensory or Intuitive, Sequential or Global.

Moreover, it was realized that Index of Learning Styles (ILS) Questionnaire [3] [2] exists to determine the dimension that an individual fit into, from the aforementioned set of learner dimensions[4], devised by Richard M. Felder and Barbara A. Soloman [4]. The Questionnaire is lengthy and consists 44 multiple choice questions [5][4]. It is proven that longer questionnaires have lower response rates by Micheal J.Roszkowski and Andrew G.Bean [6].

Upon further investigation, Petteri Nurmi and Tei Laine states that through HCI User Modeling, we can find the goals, knowledge background, traits, context of work and also the interests of a user who interacts with a system [7]. Using an Analytical Model, a system can simulate the cognitive process that carries out while a user interacts with a system [8]. Hence it explains that given user's engagement to a specific web content can be analyzed by using on screen HCI Analytical Techniques such as mouse hover, scroll, rate to content, flip, and skip watching the content and such by User Experience of On-Screen Interaction Techniques research done in 2013[8].

Evidently, with more investigation, it is found that a two-dimensional mapping which is prepared considering learning style characteristics and MOOCs characteristics exist[3] in order to help decide, which learning material type suits which learning style., i.e. if a MOOC video contains a given percentage of a talking head video style, then the MOOC video is more likely suitable for an intuitive, verbal and global learner[3]. Therefore, based on above facts, it was concluded that based on human computer interaction techniques, we can identify how engaged that individual is, with the given task.

Users who learn through peer-to-peer interactions need to find MOOCs which have an active community, so that the user too will be able to participate in it. Previous research has shown that discussions have a positive impact on learning effectiveness [9].

The research paper [10] describes an intelligent web crawler called 'iRobot' which can identify web forums with minimal initial data and figure out how to traverse the forum. Though the system in that paper does not find content relevant to a specific MOOC or topic it has a major advantage in the fact that it can identify crawling paths through forums by itself, therefore it will be able to cover a larger number of links.

The web crawler described in [11] had aimed at gathering data using regular expressions which is a template-based processing method. As a result, the information acquired is structured which makes it easier to process them later, but the disadvantage to this method is that the regular expressions must be explicitly written for the data that the system is attempting to collect. So, the system could miss out on potential information if it is not designed to do so.

A general sentiment analyzing model which was trained using 300,000 twitter posts was presented in [12]. The results of the research paper show that the model was able to identify user sentiments with decent accuracy. A multinomial Naïve Bayes classifier had been chosen to use the model data to make the classifications.

Sentiment analysis has been done in a much different approach in [13] by B. Pang et al. Before the data is used to trained or classified, the system compresses and removes any unnecessary subject domain text while retaining polarity information. Basically, it compresses the data before processing, as a result the performance was increased. In addition to that, the process increases the accuracy because the removal of unnecessary data makes the data cleaner and has a lower chance of being misidentified by the Naïve Bayes classifier.

Though previous research has shown positive results in the field of web crawling and sentiment analysis, they were not regarding the context of discussion forums relating to MOOCs. Therefore, research is needed to determine whether information (ratings derived from sentiments, other forum thread metadata) contained within forums can be used to improve MOOC recommendations.

MOOC resources include multiple modalities such as lecture videos, audio transcriptions, slides, textbooks, forum discussions and clickstream log data. Among them, lecture video is arguably the central and omnipresent component for knowledge transfer, to which other data modalities support [14]. Thus, we focus on designing a method that can organize video resources to dynamically fit different learners. Generally, the types of content in MOOC videos include talking head, slide, coding, animations, writing (khan academy), conversations etc. In production of some MOOC videos, there are lot of transitions between visual views, e.g.: switching from a talking head to a slide or switching from a slide to an animation.

1.2 Research Gap

Even though MOOCRec V1 has introduced a set of key features that were not present in any other tool a year ago, through literature review we identified a set of key areas that can be improved or extended along with a new feature that is not present in MOOCRec V1. Also, during the phase of literature review, there were no other tools that matched the feature set offered by MOOCRec V1; which led us to compare MOOCRec V1 and our solution, which is referred to as MOOCRec V2.

In areas such as processing, using resources optimally is an area that MOOCRec V1 has not covered. Also, it was clear that newer video styles can be introduced to broaden the number of recommendations that can be provided across a range of learning dimensions. Furthermore, how the learning-dimension/style of a user is identified is another aspect that could be improved. Lastly, by adding forum analysis to strengthen how MOOCs are recommended to users, we will be improving, and extending the user-experience, performance and functionality in our solution, MOOCRec V2.

As a result, the following feature comparison can be highlighted.

Table 1.1: Feature comparison between existing products vs solution

Features	Class Central	My MOOC	MOOCRec	MOOCRec V2
Directly identify users' preferred MOOC learning material type(s)	✗	✗	✗ *based on learner style*	✓
Video Production Style of MOOC videos	✗	✗	✓	✓
Complex and mixed video production styles of MOOC videos	✗	✗	✗	✓
Identify the spoken language of the presenter in a MOOC video	✗	✗	✗	✓
Search filter based on specific keywords / topics	✗	✗	✓	✓

User profile and dashboard	✓	✓	✓	✓
Online discussion forums analysis and extraction of sentiments of forum posts for better MOOC recommendations	✗	✗	✗	✓

1.3 Research Problem

As more and more people across the world adopt e-learning platforms to increase their knowledge and gain certifications, Massive Open Online Courses have come to the spotlight for their openness and scale. Along with that, search engines and review sites solely focusing on MOOCs emerged to help people find MOOCs that belong their field of study easily.

However, the clear majority of services that offer MOOCs like Coursera, Edx and search engine platforms like Class Central only take a person's interest in a certain field of study or a certification into account when filtering or recommending MOOCs. However, a service called MOOCRec has done a great deal to bridge this gap between finding a MOOC that not only matches a person's field of study but also their personal learning style. Furthermore, MOOCs belong to a wide range of video styles in which they are delivered such as animations, presentation slides, etc. This present a dilemma to a consumer as to which MOOC to choose since some video styles may appeal to a consumer more than the other.

While MOOCRec achieves many aspects of providing personalized recommendations based on individual learning styles, we have found a set of areas where the approaches can be further improved while introducing a new set of features to solidify the recommendation process by factoring in sentiments of MOOC consumers. To elaborate further, we identified that we can improve upon the process of identifying the individual learning style of a person which is done in terms of a lengthy Questionnaire in the existing product, MOOCRec V1.

Therefore, our research problem revolves around coming up with a more precise and elegant way to identify a person's individual learning style(s) and cover a broad range of MOOC video styles in order to increase the number of individual learning styles that can be supported by MOOCRec V2 service.

1.4 Research Objectives

1.4.1 General Objectives

MOOCRec V2 solution is a research based implementation of a MOOC recommender that builds up upon MOOCRec V1 but extends its capabilities while streamlining the user experience, introducing new features that capture MOOC consumers' sentiments on discussion forums and identify a broader range of MOOC video styles while moving to a new architecture that focuses more on parallel processing to speed up the process of analyzing and recommending MOOCs. To elaborate further, the objectives of this research can be outlined as follows.

- Help users identify their individual learning style(s) in an interactive, engaging, video-based tutorial that is short and specific.
- Recognize Machine Learning techniques applicable for scenarios that call for it and explore how they can be implemented to support the parallel processing architecture.
- Explore image processing and image recognition patterns and techniques.
- Integrate MOOC related online discussion spaces and ratings to solidify the recommendations made for each user.
- Index MOOCs that are presented in newer and complex video styles to create a broad spectrum of selections for video recommendations.

1.4.2 Specific Objectives

To achieve aforementioned general objectives, we have identified a range of specific objectives that have to be achieved in order to bring our solution, which is MOOCRec V1 to fruition.

- Classify and index MOOCs that are delivered in video styles that are not covered by MOOCRec V1 such as
    - Animations
    - MOOCs with multiple video styles
- Decrease or eliminate false-positives that occur with certain styles of videos such as
    - False-positives involved with an image of a person being mistaken as a talking-head video.
    - False-positives involved with predicting that a video has code in it for text that is not code.
- Meta-data in forums within the MOOC platforms and from external discussion forums like Reddit will be analyzed to give a 'forum score' which can be used to recommend MOOCs especially to users who learn through peer-to-peer interactions.
- Users' sentiments in forums within the platforms and from rating systems by services like Class Central where users have discussed about the course will be used to calculate a 'course score' which can be used when recommending MOOCs in addition to the other factors which are taken into account (learning styles, video type, etc.) to all users in general.
- Create an engaging, introductory video that contains multiple MOOC learning material types throughout its runtime to be used to capture user's engagement during each learning material type by using HCI techniques and decision making algorithms.

- Devise an architecture to classify videos parallelly in distributed computing environments to speed up the classification process and implement it.
- Provide basic user-profile related features and a dashboard to that each user can view their personalized recommendations.

2 METHODOLOGY

2.1 Methodology

The overall system is exposed to the user as a RESTful API to the users via a web interface while the backend of the system itself consists of the API and Python services that process data, as depicted in figure 2.1.

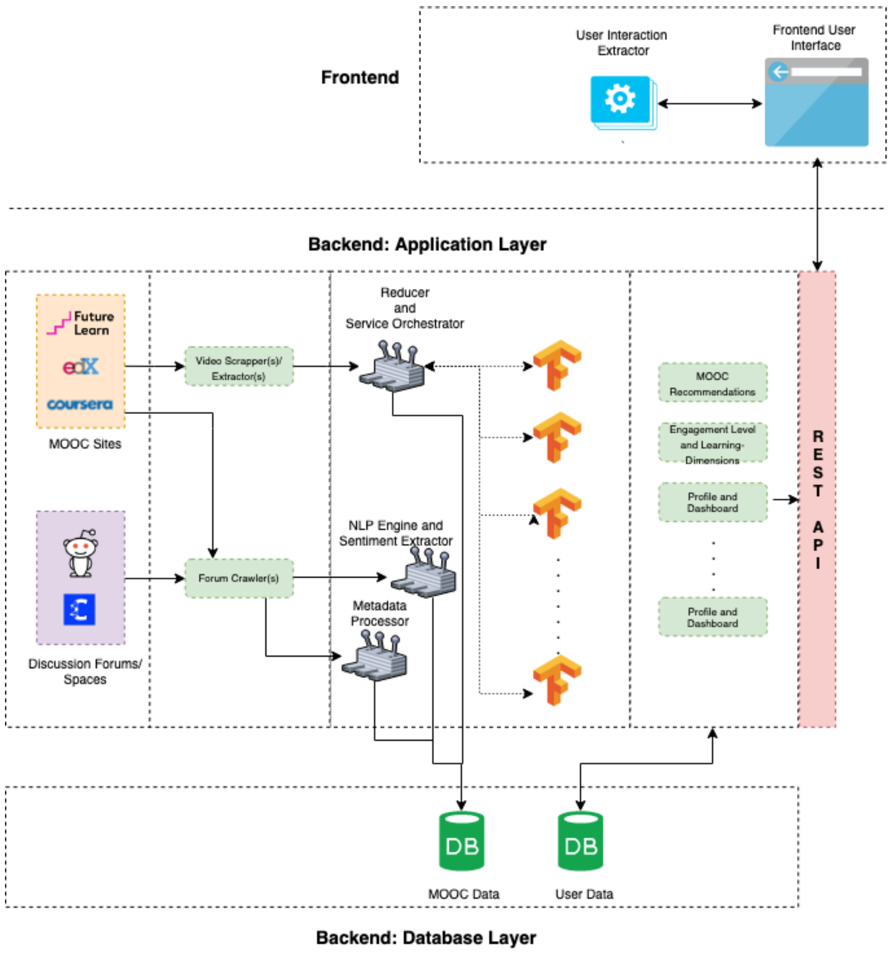


Figure 2.1: MOOCRec v2 System Diagram

MOOCRec V2 system extends upon an already existing product which has been referred to as MOOCRec as well as MOOCRec V1 in order to distinguish our system as MOOCRec V2. Our system relies on HCI techniques to identify how engaged the user is with respect to different styles of videos to score the engagement level. This highlights the connection between a certain set of user interactions and the amount of engagement that a specific interaction can imply. Therefore, by detecting the HCI of a user by analyzing browser activity, we can effectively determine the engagement level. MOOCs from popular MOOC sites such as Coursera, Edx, Futurelearn that fall into Information Technology, Science will be classified into different video styles that each video is delivered in.

The classification process is done parallelly by using multiple instances of the same classifier. These classifier-instances will be fed different frames from different videos while a service-orchestrator keeps track of which frame belongs to which video and at what time of the runtime. By doing so, the system becomes highly scalable due to its nature of independently classifying frames without any regard to where the frame is in the video.

Also, online discussion spaces and rating/review platforms offered by MOOC sites, MOOC search engines as well as other independent forums like Reddit will be periodically analyzed to extract information on MOOCs. While a web-crawler designed for each space/forum is responsible for extracting text, a suitably trained classifier will identify the sentiments such as whether the users prefer the MOOC that they are talking about or not and the amount of forum activity for a MOOC will be identified so that users who prefer to learn through peer-to-peer interaction can be suggested to them.

Finally, the data made by classifier-instances will be put together by the service-orchestrator to provide a meaningful conclusion about each MOOC that the system indexed; And the data from Online Discussion Sentiment Extractor and the Facial Expression Analyzer's conclusion on what the user's individual learning style(s) will be used to provide recommendations to each user to suite their individual learning style(s) and fields of studies.

Given its use of Machine Learning and parallel processing to achieve the implementation mentioned above, the system can be evaluated in following aspects.

- Accuracy of user-engagement evaluation.
    - Give that the system uses facial expressions and a model based on facial expressions and engagement to predict the user's engagement with a specific learning material type, the accuracy of the prediction is, by far, the most important metric to evaluate. Although the Facial Expression Analyzer initially provides the engagement level per learning material type over the course of a video that contains multiple styles, we later map the highly-engaged learning material type(s) with learning style(s). Therefore, this can be evaluated in two ways;
    - By letting a set of people freely choose their preferred learning material type(s) out of a list that contains the learning material type covered by our system, and then by analyzing the engagement-level predictions made for each user, we can identify if the system predicated high levels of engagement for the same learning material type(s) that the user freely selected initially.
    - By letting the same set of people fill the ILS questionnaire to identify their learning style and allow them to interact with our intro video and asking them what learning material type(s) they most liked we can validate our MOOC learning material-to-FSLSM mapping.
- Accuracy of classification.

  By feeding a set of evaluation/test data that consist of frames that depict different video styles from different MOOC sources, and comparing the predictions made by the classifier to the labels we provided along with the test data as to what each frame should be, we can measure the accuracy of the classifier.

- Speed.

Given the use of parallel execution and service-orchestration to classify MOOCs, we can measure the not only the speed of classification but also how much has improved over sequential classification. We can use a single video clip, and measure the time taken to classify it on a single thread using a single instance of the classifier and the time taken to classify the same video parallelly using multiple instances of the classifier to find the improvement and the speed.

2.1.1 Identifying user's preferred learning material type(s)

To identify user's learning style in a more practical manner, MOOCRec V2 is replacing the ILS questionnaire which helped user identify the learning style with an introductory video which will be identifying user's preference by tracking Human Computer Interaction. In 'Identifying user's preferred learning material type of a MOOCs user through HCI techniques phase, the system displays an interactive introductory video session which consists of all MOOCs learning material types that are recognized by the proposed system namely animations, talking head, presentation slides, khan academy writing, code/tutorial that are suitable for all types of learners. These videos are of same duration and of same subject weight. [Figure 2.2]

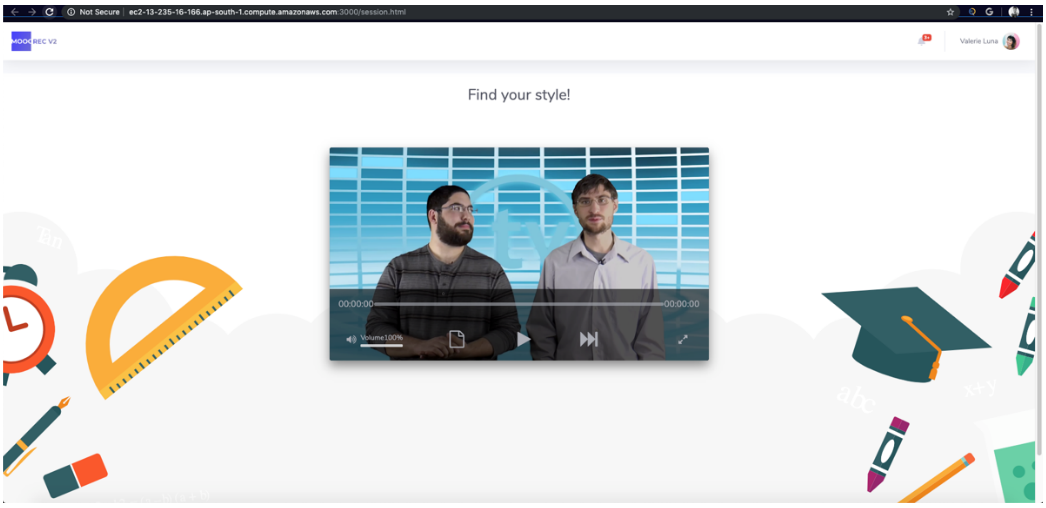


Figure 2.2: Intro Video

Our HTML 5 video player is designed in a way where users can play, pause, skip to next video, replay the same video, skim through video as well as rate the video content. Furthermore, the HTML player is designed adopting HCI techniques such as affordance, constraints, attention and workload models. During this session, we collect user feedback by using analytical HCI techniques such as dropdown point analysis, skim through rate, getting user ratings, mouse scroll motion captures (for transcript viewing), button clicks, etc. [Figure 2.2]

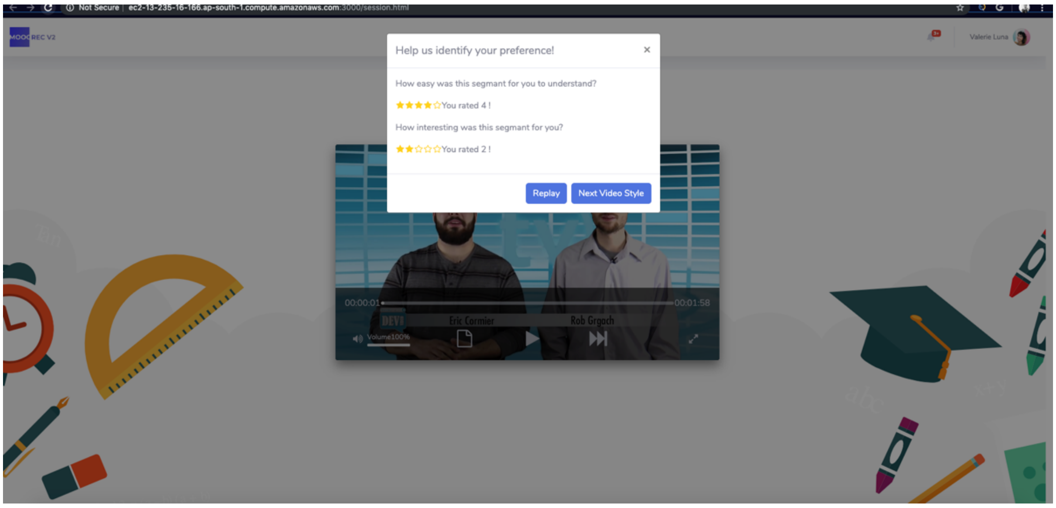


Figure 2.3: HCI Technique Rating

Moreover, each user interface related to video session has to be verified using HCI techniques to see whether following goals are achieved.

- User is aware of the interactive elements
- Interactive elements represent the input needed from the user clearly
- Interactive area is smooth, and responsive
- Interactive area adapts to the video style that is playing at the moment

As a result, our algorithm determines, during which phase of the interactive video session, the user was mostly engaged with the system. By doing so, we are left with a sizable amount of information about the engagement of the user across multiple video production styles that belong to different learning dimensions. Using this information, the application predicts the video production style(s) that is in tune with the user's learning dimension. To determine the corresponding learning style for a given video style, we carry out a mapping of MOOCs with

FSLSM (a video style to learning style mapping) which was derived through literature review [Table 3].

Table 2.1: Learning-Dimension to Video Style Mapping based on FSLM

	Talking Head	Animation	Code/ Tutorial	Presentation slides	Khan academy writing
Sensory			✓	✓	✓
Intuitive	✓	✓			
Visual		✓		✓	✓
Verbal	✓		✓	✓	✓
Active		✓	✓		
Reflective				✓	
Sequential			✓	✓	
Global	✓	✓			

Finally, the preferred learning material type prediction result is derived through a decision- making algorithm.

Consider a single segment(i) in the interactive session.

$$A = [\ \Delta ti\ T * (-1)\ ]$$

$$where\ \Delta ti\ is\ the\ skipped\ duration\ of\ segment\ i\ and\ T\ is\ the\ total\ lenght\ of\ that\ segment$$

$$B = [\ \Sigma(Pn) * 1\ ]$$

$$where\ Pn\ represent\ a\ positive\ action\ in\ segment\ i$$

$$C = [\ \Sigma(Nn\ ) * (-1)\ ]$$

$$where\ Pn\ represent\ a\ positive\ action\ in\ segment\ i\ ($$

$$D = Q1 + Q2$$

$$where\ Q\ represent\ a\ question\ asked\ in\ the\ segment\ i$$

$$\boldsymbol{Score}\ (\boldsymbol{Si}\ ) = \boldsymbol{A} + \boldsymbol{B} + \boldsymbol{C} + \boldsymbol{D}\ (10)$$

By adding the positive points gained by positive interactions and the ratings provided by the consumer at the end of each segment, as well as the negative points gained by skipping, seeking through the segments, we can give a final score to each segment that represent how positively engaged, the consumer was, with the session. [Figure 2.3]

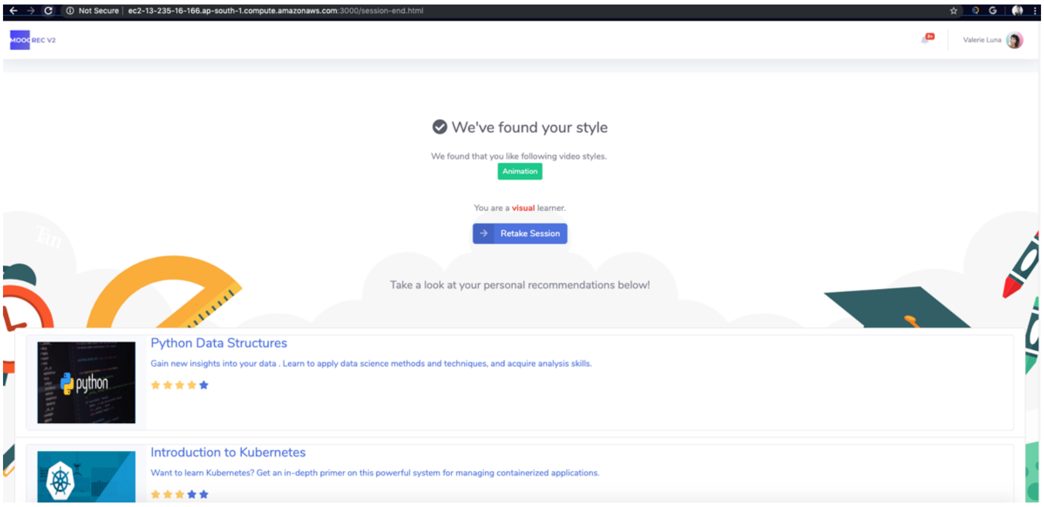


Figure 2.4: Intro Video Results

This whole process of identifying users' preferred learning material type component is graphically illustrated in [Figure 2.5].

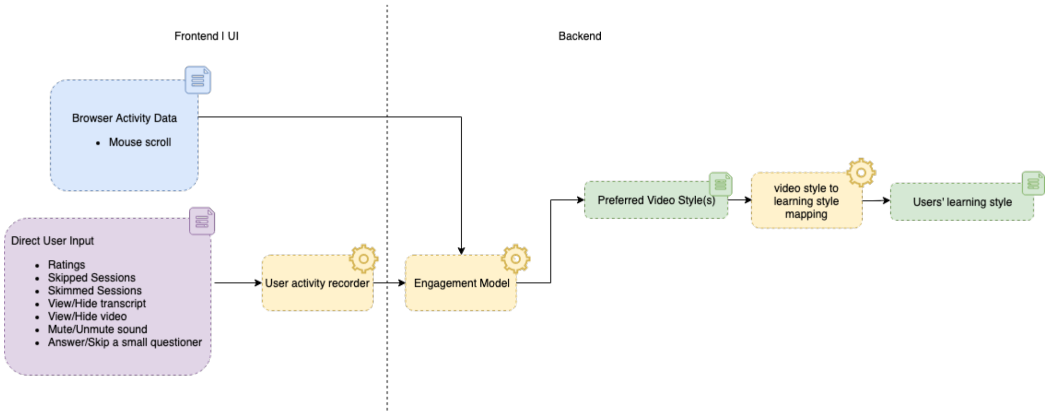


Figure 2.5: Identifying Users' Preferred Learning Type System Diagram

2.1.2 Complex Video Classifier

MoocRec1 has classified only coding, slides and talking head styles. In moocRec2, we are classifying all 6 types of video production styles (coding, slides, talking head, animation, khan academy writing and conversions). Transfer learning has been used to train the VGG16 model used by MoocRec1. Transfer learning generally refers to a process where a model trained on one problem is used in some way on a second related problem. In deep learning, transfer learning is a technique whereby a neural network model is first trained on a problem similar to the problem that is being solved. One or more layers from the trained model are used in a new model trained on the problem of interest[15].

OpenCV has been used to split video frames and image into frames. Then image frames are classified into video styles and composition of each style is calculated. For this, image-based classification approach has been used. Each frame of a video is classified into one of the video production styles. The composition of each video style is obtained by calculating the average of each style in a single video and then the average values for each MOOC course is calculated.

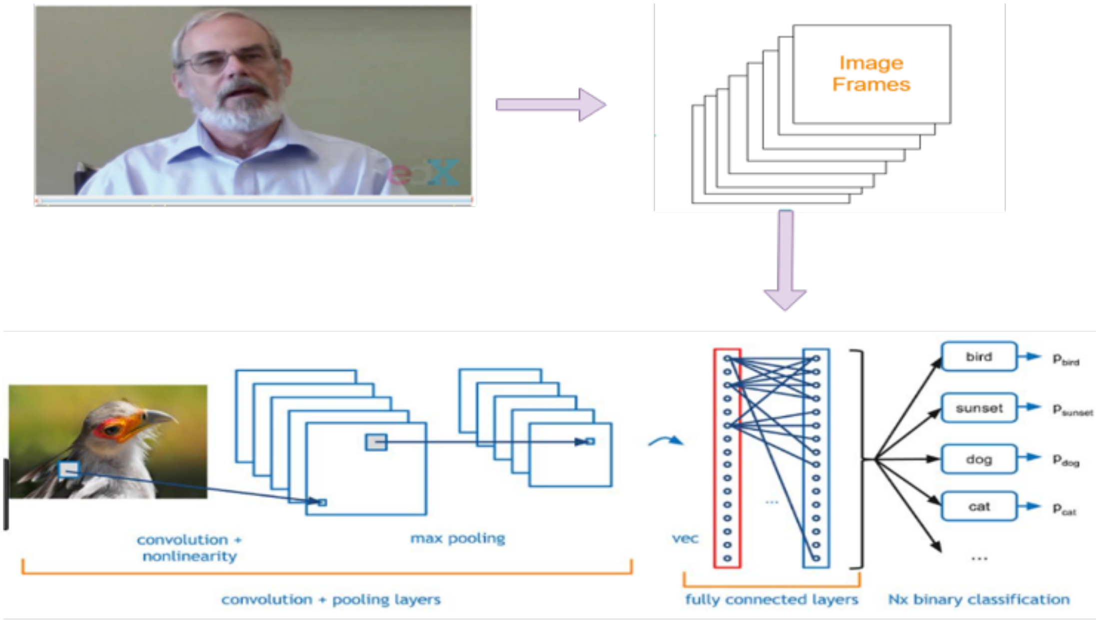


Figure 6 Fragmentation and classification of a video

If n is the total number of frames split by the classifier and a is the number of frames classified as animations, then the composition of animation in a single video is given by:

$$animation = a / n * 100$$

Finally, total composition of the video of the video styles of a course is calculated by calculating the average of each video production style.

2.1.3 Online forum thread discussion analysis to aid in recommending MOOCs

The forum analysis process requires a significant amount of data to be able to calculate the ratings. The larger the amount of data the more accurate the ratings tend to be as outliers are balanced out.

To satisfy this, three web crawlers were implemented to gather data from the MOOC platforms Coursera, Edx and FutureLearn. The Coursera and FutureLearn crawlers gather data by simulating a web browser and extracting data from the rendered web pages. The Edx crawler uses a hybrid approach as it simulates a web browser to authenticate and navigation purposes but collects the HTTP responses while they are on their way to the browser. The Edx crawler is capable of gathering more in-depth information than its counterparts due to the fact that it extracts JSON data which are received directly from the server.

Once the data gathering processes have concluded the analysis process will begin. The result of the analysis process is two rating scores, the overall course rating score and the forum activity score. The overall course rating is a measure of how useful the learner the course is while the forum activity score is a measure of how active a forum is. The high-level diagram of the forum analysis component is shown in Figure XX.

Online discussion forums regarding MOOCs generally fall into two categories; questions or discussions. Questions are threads in which discourse about the course content takes place. On the other hand, discussion-type threads are regarding the MOOC itself i.e. how well the teacher explains the topic, the language fluency, whether the topic of the subject is covered, etc. Complaints usually fall into this category and will be detected by the sentiment scores of the thread text data as it will show up as having relatively lower scores. Sentiment of text refer to how positive or negative the piece of text is.

The course rating is calculated based on the course rating available on the MOOC platform the course was extracted from and the sentiment scores of discussion-type threads as shown in equation 13.

$$\mathrm{C} = (K_1 * \mathrm{P}) + (K_2 * \mathrm{S} \ )$$

Where,

C = Course Rating

P = The rating the course has on its platform

S = The sentiment score calculated based on the discussions

K_1 and K_2 are constants of values between 0 and 1

The forum rating is calculated based on the statistics derived from the metadata of the forum threads. Metadata here refers to the attributes of the threads and posts such as 'thread creation date', 'no. of unique users on thread', 'no. of replies(posts) of a thread', etc.

2.1.4 Service Orchestrator for Parallel Classification Video Classification

The main focus of this research is classification of videos parallelly. Therefore, a single video file that has to be classified can be represented as a typical workload. Hence, breaking down a single workload involves logically marking smaller chapters of the video file. This is referred to has chunking henceforth. Chunking of a single video file is achieved as follows.

ΔT_c : $Duration\ of\ a\ chunk\ in\ seconds\ (Pre - determined)$.
ΔT : $Duration\ of\ original\ video\ file$.
f : $Frames\ per\ second\ metric\ of\ original\ video\ file$
$\sum F$: $Total\ number\ of\ frames\ of\ original\ video\ file$
$\sum F_c$: $Total\ number\ of\ frames\ in\ a\ single\ chunk$

$$f = \frac{\sum F}{\Delta T} \tag{1}$$

$$\sum F_c = f * \Delta T_c \tag{2}$$

Once number of frames that should be in a single video chunk is calculated by using equation (1) and (2), this calculated value can be used as the step value by the actual classifier when tasked with classifying a specific chunk. The figure 2.1 depicts how $\sum F_c$ value can be used to identify where each logical chunk starts and ends on the actual video file.

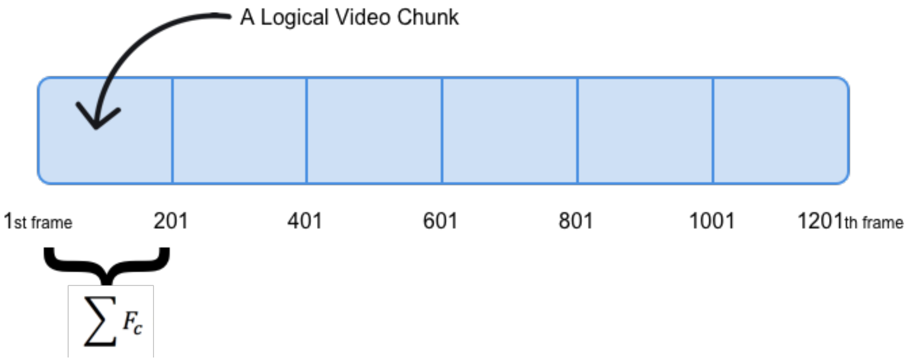


Figure 2.7: How a video file is logically chunked

Furthermore, it is important to note that the video file is not physically split since that will introduce another overhead to the overall process. Instead, $\sum F_c$ value is used as a step value for the classifier to identify the start frame and the end frame of a chunk as shown in equation (3).

$Frame_{start}$: $Starting\ frame$

$Frame_{end}$: $Ending\ frame$

Using equation (2):

$$Frame_{end} = Frame_{start+\sum F_c} \tag{3}$$

2.2 Commercialization Aspects of the Product

MOOCRec V2 can be pitched as global platform for finding personalized MOOC recommendations from all the major MOOC providers. This presents the opportunity to use targeted advertisements using service such as Google AdSense. Given that the targeted audience mainly consists of consumers who are passionate about e-learning, using targeted advertisements that promote subscriptions offered by MOOC and e-learning providers would be ideal.

Furthermore, MOOC providers can use MOOCRec V2 as a platform to promote their content, and support MOOCRec V2 via affiliate marketing.

2.3 Testing & Implementation

MOOCRec V2 was implemented as a web application for the consumers to interact with while backend was distributed among multiple nodes where video crawling, forum crawling, and engagement analyzer where given their own nodes. Video classification was distributed among multiple nodes.

The said web application was used for testing with real users to determine the ease of use, accuracy of identifying the consumer's preferred video style(s) and to determine the preference of consumers between interactive video session of MOOCRec V2 and the ILS questioner offered by MoocRec.

3 RESULTS & DISCUSSION

3.1 Results

While MOOCRec V2 is a single product, it was evaluated in multiple aspects. While the overall product is evaluated for its simplicity of use, accuracy of recommendations as well as accuracy of identifying the consumer's preference, the invidiual backend compoennts were tested for their performance and accuracy to verify that MOOCRec V2 works as a coherent and accurate product.

Test Number 01		
User	**User's preferred video style**	**System Suggestion**
User 1	Animation	Animation
User 2	Talking Head	Animation
User 3	Talking Head	Talking Head
User 4	Presentation Slides	Presentation Slides
User 5	Presentation Slides	Presentation Slides
User 6	Animation	Animation
User 7	Code	Presentation Slides
User 8	Talking Head	Talking Head
User 9	Animation	Animation
User 10	Animation	Animation

Table 3.2: Test 01 - Results

Evaluation Criteria	Validate users' own preference with system's suggested preference.
HCI Evaluation Technique(s) Used	Think Aloud, Experimental Evaluation
Test group	IT Undergraduates of age 20-25 years
Number of testers	10
Success (%)	80%
Failure (%)	20%

Table 3.3: Test 02

Test Number 02				
User	**User's preferred video style**	**Learner types suggested by FSLSM to video style mapping**	**Learner types suggested by ILS questionnaire (Listed from highest to lowest score)**	**Mapping success fraction**
User 1	Animation	Intuitive Visual Global	Intuitive Global Visual Active	3/4
User 2	Animation	Intuitive	Reflective	0/4

		Visual Global	Sequential Sensory Active	
User 3	Talking Head	Intuitive Verbal Global	Visual Active Sensory Sequential	0/4
User 4	Presentation Slides	Sensory Visual Verbal Reflective Sequential	Global Visual Intuitive Active	1/4
User 5	Presentation Slides	Sensory Visual Verbal Reflective Sequential	Visual Intuitive Reflective Sensory	3/4
User 6	Animation	Intuitive Visual Global	Active Visual Sensory Global	2/4
User 7	Presentation Slides	Sensory Visual Verbal Reflective Sequential	Active Global Verbal Intuitive	1/4
User 8	Talking Head	Intuitive Verbal Global	Active Sensory Reflective Sequential	0/4
User 9	Animation	Intuitive Visual Global	Verbal Sensory Sequential Active	0/4
User 10	Animation	Intuitive Visual Global	Active Sequential Reflective Verbal	0/4

Table 3.4: Test 02 - Results

Evaluation Criteria	Validate FSLSM to Video Style Mapping.
HCI Evaluation Technique(s) Used	Think Aloud, Questionnaire
Test group	IT Undergraduates of age 20-25 years
Number of testers	10
Success (%)	25%
Failure (%)	75%

Table 3.5: Test 03

Test Number 03		
User	**Time taken to complete intro session (seconds)**	**Time taken to fill ILS questionnaire (seconds)**
User 1	312	1200
User 2	600	900
User 3	501.6	1500
User 4	571.2	1155
User 5	502.2	1261.2
User 6	455.4	1830
User 7	375	1265.4
User 8	503.6	1503.6
User 9	452.4	1080.6
User 10	367.2	1116

Table 3.6: Test 03 - Results

Evaluation Criteria	Compare the time taken to fill ILS questionnaire and the time taken for an average person to complete interactive session.
Time measurement tool	Stop watch
Test group	IT Undergraduates of age 20-25 years
Number of testers	10
Time taken to fill ILS questionnaire per tester	21.353 minutes / 1,281.18 seconds
Time taken to complete intro session per tester	7.734 minutes / 464.06 seconds

Forums analyzer was tested for its capability to extract as many as threads from a given platform as well as time taken to extract the said threads. This was done to determine how feasible and practical it is to extract forum threads to identify the sentiments of the posters.

Platform	Course	Actual No. of Threads	Extracted No. of Threads
FutureLearn	Introducing Robotics	242	242
Edx	Python for Data Science	1971	1971
Coursera	Machine Learning	134,946	134,946

Platform	Course	Time Taken to Extract Data (seconds)
FutureLearn	Introducing Robotics	77
Edx	Python for Data Science	498
Coursera	Machine Learning	1472

While table .. and table .. shows the practicality and accuracy of forum crawler, the below table shows how accurate the sentiment extraction from forum threads were between Google NLP and VADER Sentiment implementations. Each implementation was tasked with predicting emotions expressed by a statement extracted from a forum thread.

Platform	Course	Average Time Taken to Analyze all Threads of Course (second ds)	
		Google NLP	VADER Sentiment
FutureLearn	Introducing Robotics	26	20
Edx	Python for Data Science	240	108
Coursera	Machine Learning	588	342

Another major component of MOOCRec V2 is classifying MOOCs into 6 different video production styles. Below table shows the accuracy of VGG16 model in predicting the video production style of MOOC courses.

<sachin’s tables>

Furthermore, the classifier was containerized and parallelized by using a distributed compute platform. Table .. shows how efficient distributed parallel classification is compared to single thread-based classification.

Workload Type	Instance Type	vCPUs per Instance	Memory per Instance	Number of Instances	Total Compute Time Taken (Seconds)			
					Run #1	Run #2	Run #3	Average
Centralized	t2.medium	2	4GB	1	3545	4157	4183	3962
Distributed	t2.micro	1	1GB	2	2168	2107	2090	**2122**

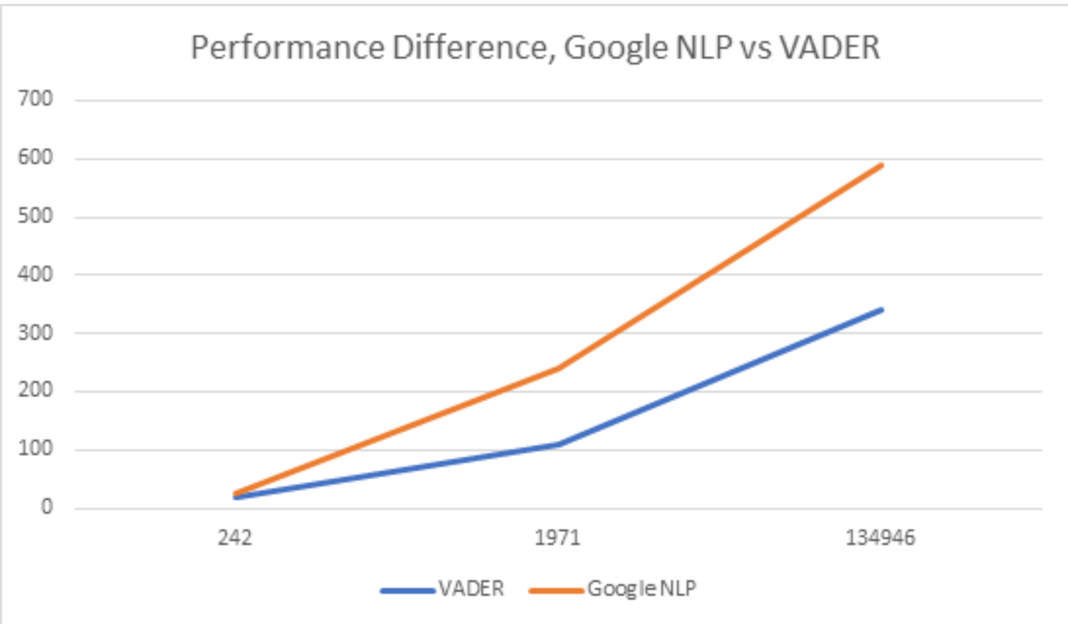
3.2 Research Findings

MOOCRec V2’s interactive video session was introduced as a direct way to identify the consumer’s preference in video styles. While using ILS questioner to identify the learning style of a consumer, which in return is used to predict the video style preference of the user is already in motion, directly identifying the preferred video style of a consumer by analyzing their engagement on different video styles appears to be the more direct approach. Furthermore, testing has shown that this direct approach yields more accurate predictions.

According to the aforementioned results which were found during the testing and evaluation phase of MOOCRec V2, it was found that the attempt of mapping FSLSM with video styles was not a massive success since the success percentage was only 25% whereas the failure percentage was 75%.

Furthermore, it was found that MOOCRec V2 is successfully able to conquer and provide 80% accurate suggestions regarding one's preferred MOOCs video style. Apart from that, the attempt of designing an introductory session to identify one's preferred video style within a shorter period of time, rather than spending a lot of time in order to complete ILS questionnaire turned out to be a success.

Different online discussion forums have different styles of threads and discussion spaces. Therefore, a dedicated crawler for each forum had to be implemented which was specialized to deal with the schematics of the forum it is associated with. This proved to be more capable than a single generalized crawler, by extracting 100% of the available threads on tested forums.



Out of many pre-trained image classification models, VGG16 was determined to yield the most accurate result in classifying videos into 6 different video production styles.

3.3 Discussion

In the forum analysis component from the two methods tried out, the VADER library is more efficient. Efficiency is important because there are a large number of MOOCs available in the three platforms. Forum data of a MOOC though not highly volatile does change over time therefore needs periodic updating. Performance is a desired characteristic in scenarios where constant updating is performed. As the data in Figure 3.1 shows VADER is significantly faster than Google NLP. Google NLP seems to be more accurate than VADER as its model is constantly improved and is a paid service, but in this case, performance is much more important than accuracy. A minor reduction in accuracy will not adversely affect the system unlike low performance.

When it comes to the video analysis parallelization component, the testing that was done to determine the performance gain that can be achieved by parallelizing an image classification workload across a distributed computing platform can be said to have yielded positive results. While this method of classifying videos and images clearly does not yield performance gains that are comparable to GPU based parallel processing, it does however yield more than 75% performance improvements for a lesser cost as well.

According to the results which were found during the testing and evaluation phase of MOOCRec V2, it was found that the attempt of mapping FSLSM with video styles was not a massive success since the success percentage was only 25% whereas the failure percentage was 75%.

Furthermore, it was found that MOOCRec V2 is successfully able to conquer and provide 80% accurate suggestions regarding one's preferred MOOCs video style. Apart from that, the attempt of designing an introductory session to identify one's preferred video style within a shorter period of time, rather than spending a lot of time in order to complete ILS questionnaire turned out to be a success.

4 SUMMARY OF EACH STUDENTS CONTRIBUTION

Student ID	Student Name	Contribution
. IT 16 0519 80	De Silva W.A.T.P	• Engagement analyzer • Interactive Video Session
IT 16 0327 98	Liyanage A.Y.K.	• Containerize classifier • Implement message queue • Devise and implement the algorithm to map reduce outputs without waiting for all the workloads to complete

IT16005372	Hilmy S.B.M	• Forum crawling • Sentiment extraction • Sentiment analysis
IT 16 0043 82	P.H.P.S.L. Pathirana	• Classify 6 video production styles • Reduce false positives •

5 CONCLUSION

Massive Open Online Courses have become a popular means of learning in a wide variety of subjects. MoocRec V1 introduced novel features to solve the problem of helping users find MOOCs. This paper proposes an improved version of MoocRec V1 by overcoming its issues and adding new features which would help users find the MOOCs that suit them the most. The results of the tests conducted show that the new features proposed are sufficiently effective at improving the system's usefulness.

6 REFERENCES

[1] M. Khalil, "Learning Analytics in Massive Open Online Courses," no. April, 2018.

[2] R. M. Felder, "A Longitudinal Study of Engineering Student Performance and Retention. IV. Instructional Methods," *J. Eng. Educ.*, vol. 84, no. 4, pp. 361–367, 1995.

[3] S. A. It, "MOOCREC : LEARNING STYLES-ORIENTED MOOCS RECOMMENDER AND SEARCH ENGINE Department of Software Engineering Sri Lanka Institute of Information Technology Sri Lanka MOOCREC : LEARNING STYLES-ORIENTED MOOCS RECOMMENDER AND SEARCH ENGINE October 2018," no. October, 2018.

[4] S. Graf, Kinshuk, and Tzu-Chien Liu, "Identifying Learning Styles in Learning Management Systems by Using Indications from Students' Behaviour," 2008.

[5] R. M. Felder and B. A. Soloman, "Index of Learning Styles Questionnaire."

[6] M. J. Roszkowski and A. G. Bean, "BELIEVE IT OR NOT ! LONGER Q U E S T I O N N A I R E S H A V E LOWER R E S P O N S E RATES," vol. 4, no. 4, pp. 495–496, 1990.

[7] T. L. Petteri Nurmi, "Presentation: Introduction to User Modeling," p. 25, 2007.

[8] S. S. Sundar, S. Bellur, J. Oh, Q. Xu, and H. Jia, "User Experience of On-Screen

Interaction Techniques: An Experimental Investigation of Clicking, Sliding, Zooming, Hovering, Dragging, and Flipping,” *Human–Computer Interact.*, vol. 29, no. 2, pp. 109–152, 2014.

[9] C. Kent, E. Laslo, and S. Rafaeli, “Interactivity in online discussions and learning outcomes,” *Comput. Educ.*, vol. 97, pp. 116–128, 2016.

[10] R. Cai, J. Yang, W. Lai, Y. Wang, and L. Zhang, “iRobot: An Intelligent Crawler for Web Forums,” *Proceeding 17th Int. Conf. World Wide Web (WWW 2008)*, pp. 447–456, 2008.

[11] Q. Gao, B. Xiao, Z. Lin, X. Chen, and B. Zhou, “A high-precision forum crawler based on vertical crawling,” *Proc. 2009 IEEE Int. Conf. Netw. Infrastruct. Digit. Content, IEEE IC-NIDC2009*, pp. 362–367, 2009.

[12] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining Microblogging Microblogging = posting small blog entries Platforms,” *Analysis*, pp. 1320–1326, 2010.

[13] B. Pang and L. Lee, “Sentiment analysis using subjectivity summarization.pdf,” 2002.

[14] X. Zhang, C. Li, S. W. Li, and V. Zue, “Automated segmentation of MOOC lectures towards customized learning,” *Proc. - IEEE 16th Int. Conf. Adv. Learn. Technol. ICALT 2016*, pp. 20–22, 2016.

[15] J. Brownlee, “How to Reuse Models for Computer Vision with Transfer Learning in Keras,” *Machine Learning Mastery*. 2019.

7 APPENDICES

[Appendix 1: Pie Chart Representation for Survey on Filling a Questionnaire]

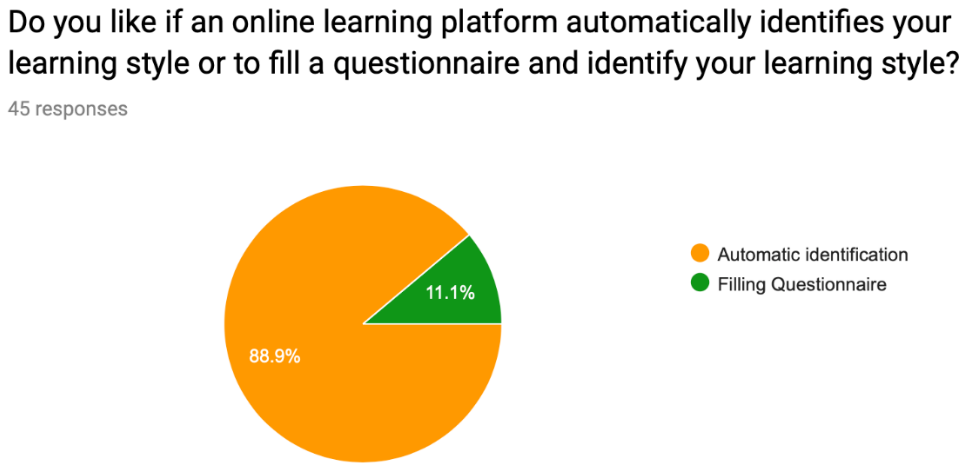


Figure 7.1: Survey on Filling a Questionnaire