



Investigating MOOCs using Social Media Mining: A Survey

Priyanka Kumari, Seema Sharma

Department of Computer Science, BIT Mesra,
India

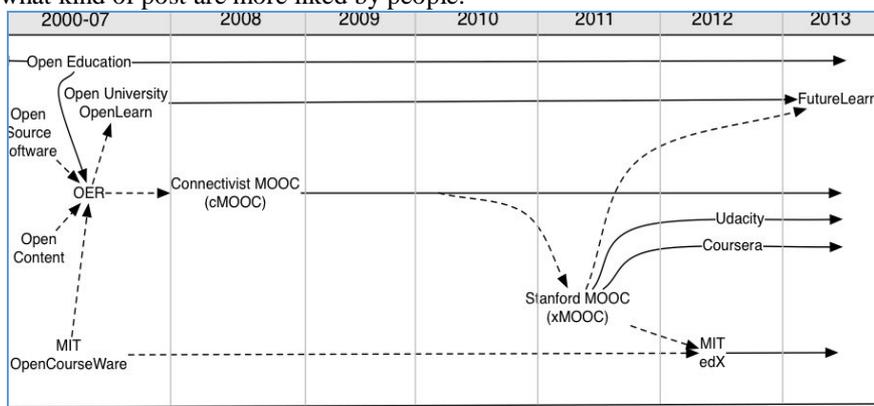
Abstract— MOOC has experienced a large expansion in the last 5 years and have gained a large popularity among the online users. Despite of its popularity a high rate of dropouts has been observed, so it's important to have a more detailed analysis of public opinion towards MOOC and its current trend to improve its performance and retention of people for a particular course. In this study we adopted various social media mining approaches to investigate Facebook post and comments related to MOOC learning. In the first approach we will show a descriptive statistic of Facebook post posted daily/Weekly/Monthly and yearly related to MOOC by a MOOC provider and user's responses through comments, like and share for a particular post. Secondly we will investigate the public opinion towards MOOC using sentiment analysis of related comments on posts. These analysis will provide us an understanding of MOOCs current trend among users and public sentiment towards MOOC learning.

Keywords— MOOC, Facebook, Social Media, Sentiment Analysis, and Data mining.

I. INTRODUCTION

MOOCs are a relatively recent online phenomenon and have come into picture since 2012. Massive Open Online Courses (MOOCs) is an online course with the option of free and open registration, a public shared curriculum and open-ended outcomes for higher education and aimed at unlimited participation and open access via the web. Some providers are Coursera, Udacity, Udemy and edX. In addition to traditional course material such as lectures, reading and problem set, many MOOC provide interactive user forums to support community interaction between students, professors and teaching assistants. MOOC are recent and widely researched development in distance education which were first introduced in 2008 and emerged as popular mode of learning in 2012. The first MOOC emerged from the open educational resources (OER) movement. According to New York Times, 2012 becomes "The year of MOOC" as several well-financed providers, associated with top universities, emerged, including Coursera, Udacity and edX. The main reason behind its popularity is that it includes accessibility to every person in the world who has internet, scalability to handle any number of student with wide diversity of needs and expectation and flexibility they provide to learners to study according to their routine.

In this paper we will analyse the Facebook data related to MOOC learning. We will extract all the data from a MOOCs provider's "Coursera", "edX" and "Udacity" Facebook page using R language. Extracted data will be used further for sentiment analysis using a tool Datum box. In first approach we will show a descriptive statistic of post, comment and share on the page and examine the daily, monthly, and yearly trends of MOOC that appeared in Facebook. These statistics can show what kind of post are more liked by people.



II. MOOC CHALLENGES

Even though MOOCs have been widely accepted, there is a lot of scope for improvement as far as the actual need of student is considered. There are many challenges which need to be considered. Issues such as high rate of attrition, lack of direct communication between instructor and learner, questionable course quality, unavailable course credit, ineffective assessment, limited hardware. Other challenges such as sustainability, Pedagogy: Do MOOCs follow a sound pedagogy and organisational approach to online learning that will lead to quality outcomes and experiences for students?.

What new pedagogies and organisational mechanism might be required if MOOC are to deliver a high quality learning experience?. Many research works have been done to improve the performance of MOOC, as the lack of direct communication is one of the major issue, many text mining approaches has been applied to understand the sentiment of students through their feedback on forums, social networking and blogs. As many MOOCs adopted social media tools for large audience and for better communication we can use those text to better understand the process which can help in increasing the performance and retention of people for a particular course. Social networking portals today has becomes a very popular communication tool among internet user. Millions of user share their opinion on different aspects of life every day. Therefore social network websites are rich source of data for opinion mining and sentiment analysis. In addition, using social media tool is central to many MOOCs because it facilitates the critical MOOC aspect of connectivity, communication and interaction. Participant can use popular social media, such as YouTube, Twitter, and Facebook throughout the courses to share content. Therefore, many MOOCs adopted social media extensively to develop ad hoc learning communities and facilitate discussions and learning. Although embedding social media in the online classroom can enhance interactivity, collaboration and student satisfaction, excessive social media messages increase the complexity of data analysis.

III. LITERATURE REVIEW

A. Methodology

The seven articles studied presented in this survey has been summarised in the Table 1. The first, second and third column describes the Title, Author and Year respectively. Fourth column describes the algorithm used for analysis, fifth and sixth column describes the data source and software application used for analysis. Last column summary describes the objective of the research done.

Chien-wen Shen, Chin Jin Kuo [2015][1] performed various social media mining approaches to investigate Twitter messages related to MOOC learning for a holistic understanding of MOOC trends, public sentiment towards MOOC learning, and the influencers of MOOC- related retweets using text mining approach on twitter data using tool Opinion finder.

First approach adopted in this study was calculating the important descriptive statistics of MOOC- related tweets and examining the daily, weekly, and monthly trends of MOOC that appeared on Twitter. Secondly the investigated how public sentiment towards MOOC learning can be assessed according to related tweets, then they analyse the positive and negative retweets related to MOOCs and identified the influencers of these retweet.

Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim [2015] [2] uses lexicon based approach to extract sentiment from forum posts and click stream logs to propose and algorithm based on artificial neural network for predicting student's attrition in MOOCs using sentiment analysis and show the significance of student sentiments in this task.

Ishan Sudeera Abeywardena [2014] [3] attempts to explore the public opinion and perceptions regarding OER, MOOC and their complementary roles. A text mining approach is used to analyse raw Twitter data in the domains of OER and MOOC within a time span of 12 months. Sentiment analysis was applied to the data to understand how public perception have changed during this time period.

Yong Chen [2014] [4] uses blog mining approach to analyse what themes and trends about MOOCs can be found. The goal of the research was to synthesize related discussions in blogs, to provide an in-depth review of MOOCs, and to identify the challenges and future trends of MOOCs. The paper hopes to aid MOOC providers and higher education institutions that might be interested in joining MOOCs to understand what is going on in this fast-moving field.

Miaomiao Wen, DiyiYang, Carolyn Penstein Rose [2014] [5] uses forum post of Courser.org for mining collective sentiment from the post in order to monitor student's trending opinions towards the course and major course tools, such as lecture and peer- assessment.

Carlos Alario-Hoyos, Mar Perez-Sanagustín [2013] [6] analyses the impact of two built-in (Q&A and forum) and three external social tools. (Facebook, Twitter and MentorMob) in a MOOC on educational technologies to reduce dropout rates. The analysis of the impact of these social tools was carried out from two different points of view: learners' perspective indicating the utilization of the social tools, and tools perspective, collecting quantitative data from their actual use. How many people Instructor or a participant uses social media actively to discuss about MOOC. What are the different way to communicate about it?

TABLE I

S. no	Title	Author	Year	Algorithm used	Data Scope	Software Application used	Summary
1	Learning in massive open online courses: Evidence from social media mining	Chien-wen Shen, , Chin-Jin Kuo	2015	Uses Naïve Bayes classifier.	Twitter	OpinionFinder	The findings pertaining social media mining in this study afford a holistic understanding of MOOC trends, public sentiment toward MOOC learning, and the influencers of MOOC-related retweets.
2	Predicting Student Attrition in MOOCs using	Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim	2015	Lexicon-based approach to extract sentiment from forum posts using	Click stream log and forum posts data from	SentiWordNet 3.0 [3]	This paper proposes an algorithm based on artificial neural network for predicting student attrition in

	Sentiment Analysis and Neural Networks	Samsung Electronics Co., Ltd. Seoul, South Korea		SentiWordNet 3.0 [3] as the knowledge resource.	Coursera		MOOCs using sentiment analysis and show the Significance of student sentiments in this task.
3	Public Opinion on OER and MOOC: A Sentiment Analysis of Twitter Data	Ishan Sudeera Abeywardena, School of Science and Technology, Wawasan Open University, Penang, Malaysia	2014	Text mining approach.	Twitter- 12 months data	Sementria-plugin for MS Excel	Chronological view of public opinion on OER and MOOC. - Search term OEM and MOOC were used.
4	Investigating MOOCs Through Blog Mining	Yong Chen, Old Dominion University, USA	2014	Blog mining	Blog data - 2010 to 2013	Google blog search, concept analysis and mapping (CAAM) technique - Leximancer	Used blog mining to analyse MOOCs to give solution to analyse what themes and trends about MOOCs can be found and to identify the challenges and future trends of MOOCs.
5	Sentiment Analysis in MOOC Discussion Forums: What does it tell us?	Miaomiao Wen, Diyi Yang, Carolyn Penstein Rosé, School of Computer Science Carnegie Mellon University 5000 Forbes, Pittsburgh, PA	2014	Brown clustering, survival modelling Technique.	From forum post of Coursera.org		Mining collective sentiment from forum posts in a Massive Open Online Course (MOOC) in order to monitor students' trending opinions towards the course And major course tools, such as lecture and peer-assessment.
6	Analysing the Impact of Built-In and External Social Tools in a MOOC on Educational Technologies	Carlos Alario-Hoyos, Mar Pérez-Sanagustín, Department of Telematics Engineering, Universidad Carlos III de Madrid Av. Universidad, 30, 28911, Leganés (Madrid), Spain.	2013	Text mining approach.	Survey	Survey	Analyses the impact of two built-in (Q&A and forum) and three external social tools (Facebook, Twitter and MentorMob) in a MOOC on educational technologies To reduce dropout rates. The analysis of the impact of these social tools was carried out from two different points of view: learners' perspective indicating the utilization of the social tools, and tools perspective, Collecting quantitative data from their actual use. How many people Instructor or a participant uses social media actively to discuss about MOOC. What are the 7different way to communicate about it?
7	Exploring the Use of MOOC Discussion Forums	D.F.O.Onah, J.E.Sinclair and R.Boyatt, The university of Warwick, United Kingdom	2012			Forum Data	Investigated the pedagogy for effective use of MOOC forum and to explore aspects of adaptivity for MOOCs.

B. Tools available for sentiment analysis.

Different tools are available in the market for the purpose of analysing the public opinion using sentiment analysis. In this research paper Datum box API has been used for this purpose as it is an open source. A comparison among the various tools is depicted in Table III.

Table III

Features	Sentiment Analysis Tools					
	R	WEKA	SentiWordNet	Sementria	Datum Box	Opinion Finder
Developed By	Ross Ihaka and Robert Gentleman [1993]	University of Waikato [2011]		Oleg Rogynskyy [2011]	Vasilis Vryniotis[2014]	O'connor et al (ICWSM 2010, Opinion Polls), Johan Bollen (2010)
Open	YES	YES	YES	NO	YES	NO

Source						
Advantage	Library support, Visualization	Free availability under the GNU General Public License. Portability, since it is fully implemented in the java programming language and thus runs on almost any modern computing platform. A comprehensive collection of data pre-processing and modelling technique. Ease of use due to its graphical user interfaces.	Uses lexicon based approach.	Excel plugin is available which is easy to use.	Uses powerful and open – source Datum box Machine Learning Framework. It uses REST and JSON technologies, offer a common interface across different classifier. Uses feature selection algorithm	Positive vs. Negative mood from text content
Official Website	http://www.r-project.org/	www.cs.waikato.nz/~ml/weka/	Sentiwordnet.isti.cnr.it/	https://www.lexalytics.com	www.datumbox.com	http://www.cp.pitt.edu/mpqa/opinionfinderrelease/

IV. CONCLUSIONS

In this survey paper we investigated how social media mining such as Facebook data mining will help MOOCs providers to find the current trends and public opinion towards MOOCs. The positive and negative sentiments of the user's comments on MOOCs related post will help MOOCs providers to investigate negative feedbacks and find the reason where MOOCs are lacking and the underlying reason for increased drop outs rate. The descriptive statistics of MOOCs related post will indicate the frequency of MOOC-related discussion on Facebook and its popularity among the users.

REFERENCES

- [1] Abeywardena, I. S. (2014). *Public opinion on OER and MOOC: A sentiment analysis of twitter data*. Paper presented at the international conference on open and flexible education (ICOFE 2014). Hong Kong, China. 2014
- [2] Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim (2015). *Predicting Student Attrition in MOOCs using Sentiment Analysis and Neural Networks*. Seoul, South Korea. 2015
- [3] Yong Chen. (2014) *Investigating MOOCs Through Blog Mining*. , Old Dominion University, USA. 2014
- [4] Miaomiao Wen, Diyi Yang, Carolyn Penstein Rosé (2014). *Sentiment Analysis in MOOC Discussion Forums: What does it tell us?* School of Computer Science Carnegie Mellon University 5000 Forbes, Pittsburgh, PA. 2014
- [5] Carlos Alario-Hoyos, Mar Pérez-Sanagustín (2013) *Analyzing the Impact of Built-In and External Social Tools in a MOOC on Educational Technologies*. Department of Telematics Engineering, Universidad Carlos III de Madrid Av. Universidad, 30, 28911, Leganés (Madrid), Spain 2013
- [6] Devendra Singh Chaplot, Eunhee Rhim. *SAP: Student Attrition Predictor*. Seoul, South Korea
- [7] Johan Bollen, Huina Mao, Alberto Pepe (2011): *Modelling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena*. Harvard University.2011
- [8] Keith Harman, Alex Koohang, (2013). *MOOC 2050: A FUTURISTIC TOUR*. Middle Georgia. 2013
- [9] Ming Liu*, Weiwei Xu, Huiran Liu, Jiaqin Zheng (2015), *Sentiment Analysis in Chinese Web Discussion Forums*. Chongqing, China.
- [10] D.F.O.Onah*, J.E.Sinclair and R.Boyatt (2014), *Exploring the Use of MOOC Discussion Forums*. The University of Warwick, United Kingdom.
- [11] Scott Crossley, Danielle S., McNamara, Ryan Baker, *Language to Completion: Success in an Educational Data Mining Massive Open Online Class*. Georgia State U.