Understanding Learners' Opinion about Participation Certificates in Online Courses using Topic Modeling

Gaurav Nanda[a] Nathan M. Hicks[a] David R. Waller[a] Dan Goldwasser[b] Kerrie A. Douglas[a]
[a] School of Engineering Education, [b] Department of Computer Science, Purdue University, West Lafayette, USA

ABSTRACT

This study proposes a formal multi-step methodology for qualitative assessment of topic modeling results in the context of online learner motivation to purchase Statements of Participation (SoP). We developed Latent Dirichlet Allocation (LDA) based topic models on open-ended responses of three post-course survey questions from 280 open courses offered on the FutureLearn learning platform. For qualitative assessment, we first determined the theme of the topic based on the words that constituted the topic and responses that were most strongly associated with the topic. Then, we verified the theme by comparing the topics assigned by LDA model on a test set with manual annotation. We also performed sentiment analysis to check for alignment with human judgment. Learner motivations in each theme were interpreted with the Expectancy-Value-Cost framework. Our analyses indicated that, primarily, learners were motivated to purchase the SoP based on perceptions of the utility value and financial cost of the certificate. We found that human judgment agreed with the topic model more frequently when LDA topic weights were larger.

Keywords

MOOC Certificates, Topic Modeling, Latent Dirichlet Allocation, Text Mining

1. INTRODUCTION

Open-ended survey responses contain rich information that is often hard to capture through closed-ended questions. Open-ended questions allow users to not only answer the question asked but also express their opinions freely, offer insights that may be novel, and provide suggestions for improvement. For an evolving system such as Massive Open Online Courses (MOOCs), where there is a large variation in the learners' backgrounds and learning objectives, it is challenging to design closed-ended surveys with predetermined options encompassing all aspects. Therefore, use of open-ended surveys that allow obtaining detailed feedback and insights from users on different aspects can be very useful. However, manually analyzing open-ended survey responses from large, diverse populations can be challenging. Data mining techniques can be helpful in this regard, but they involve issues related to interpretability of their results.

In the context of our research, the primary issue is the extent to which topics identified by topic modeling techniques represent qualitatively meaningful themes.

1.1 Topic Models

While manual analysis of open-ended responses is extremely tedious, topic modeling algorithms can find emerging themes from a large collection of documents [1] and have been used for exploratory analysis of large textual collections such as MOOC discussion forums [2]. In this study, we used Latent Dirichlet Allocation (LDA) based topic modeling, which is a probabilistic unsupervised classification method that models each document as a mixture of underlying topics and each topic as a collection of related words. The LDA model tries to identify these topics iteratively based on the co-occurrence of words in documents and represents each document as a composition of different topics with associated weights. A good explanation of the algorithm can be found in [3]. "Topic models provide useful descriptive statistics for a collection, which facilitates tasks like browsing, searching, and assessing document similarity" [4].

Notably, the topic model algorithms have no domain knowledge and the documents are not annotated with topics or keywords. However, the generated topics often resemble the thematic structure of the document collection and topic annotations by model are useful for tasks such as classification and data exploration. "In this way, topic modeling provides an algorithmic solution to managing, organizing, and annotating large archives of texts" [5].

Since topic modeling is an unsupervised method, the ground truth set of topics is unknown—which makes it hard to judge the quality and relevance of topics identified by models such as LDA. Also, the interpretability of the topics generated from these models is not guaranteed [6]. Measures such as Perplexity or Probability of held-out documents [7] have been proposed for evaluating the quality of topic models but they have not been found to correlate well with human judgment because they do not capture topic coherence or semantic interpretability [8], [9]. On the other hand, 'Topic Coherence' measures have been found to better correlate with human judgment [6], [10], [11]. Finding out the exact meanings of the topics requires additional information and domain knowledge [12]. In a study comparing human evaluation of topics with these traditional metrics, authors recommended that "practitioners developing topic models should thus focus on evaluations that depend on real-world task performance, rather than optimizing likelihood-based measures" [8]. Therefore, in this study, we conducted qualitative analysis of topics identified by LDA model to determine their theme and relevance in context of online course certificates.

1.2 Online Participation Certificates

MOOCs provide the opportunity to deliver knowledge and skills to learners anywhere in the world, at relatively low cost. Learners can document their MOOC achievements through certificates, which are increasingly becoming an acceptable medium for skill or knowledge validation among employers [13], [14]. It has also been found that learners who opt for certification in MOOCs are more likely to actively participate in and complete courses [14],

[15]. As such, identifying factors associated with certificate purchasing can lead to better participation and learning.

Our aim in this study was to understand the value that learners associate with the course participation certificate. To our knowledge, previous literature has not studied large-scale learner feedback to assess the importance of online learning certificates. In this study, we analyzed the open-ended responses to post-course survey questions from about 280 courses offered on the FutureLearn platform to understand the reasons why learners were interested or not interested in the Statement of Participation (SoP), and what would make it more appealing to them. On the platform used for this study, the SoP can be purchased by learners if they "mark over 50% of the steps on a course as complete and attempt all test questions" [16].

1.3 Learner Motivation and the Expectancy-Value-Cost Model

The Expectancy-Value-Cost (EVC) model of motivation has been shown to capture the important features of learning, persistence, and performance-based behaviors. EVC theory characterizes motivation to engage in a given task by the expectation of success, the perceived value, and the perceived cost of engaging in the task [17]. Expectancy is related to a learner's self-conception of their ability, task difficulty, and academic mindset, and it helps predict achievement. Value is based on intrinsic motivation, perceived utility, and attainment (affirmation of identity), and it is highly related to continued interest and persistence. Cost has four associated elements related to task effort, outside effort, loss of valued alternatives (including money), and emotion. Cost negatively affects both expectancy and value in different ways [18]. Because retention in MOOCs is a common problem, this study aims to understand the values and costs associated with SoPs which can help increase MOOC completion rates.

When learners decide to participate in MOOCs, they come with a wide variety of backgrounds and motivations. Their varying circumstances affect their ability to invest time, effort, and money to participate, and through EVC theory these variations can help develop our understanding and strategies to increase motivation, such as offering the chance to invest in a SoP [19], [20]. However, there are a variety of influences on learners' decisions to purchase SoPs. When a learner enrolls in a MOOC and purchases the SoP, their investment is often associated with its value and cost and can provide a motivational tool for learning and course completion. Thus, the reasons why learners do or do not purchase SoPs can inform this motivational strategy for improved retention and learning.

2. METHOD

We analyzed following three post-course survey questions:

Q1. Why are you interested in a SoP?
Q2. If no (not interested in SoP), why not?
Q3. What would make a SoP more appealing to you?

The post-course survey data was provided to us by the platform in the form of separate CSV files for each course. We first collated together all the responses to each of the listed questions from different courses. From the collected responses, we removed the records that did not contain any text. It is to be noted that considerably more learners answered the post-course survey question Q2- why they were not interested in the SoP (~56,000), than Q1-why they were interested in it (~12,600). It was encouraging that a lot of learners (~49,000) answered Q3-what would make the SoP more appealing to them. Regarding the length of responses, about 30% of responses for Q1 and Q2, and 40% for Q3, had 5 or fewer words. For all questions, about 60% responses had 10 or fewer words and about 75% responses had 15 or fewer words. For each question, we randomly selected 100 responses to be used as the TEST set and the remainder to be the TRAIN set.

2.1 Topic Modeling

We used the MALLET library [21] for developing the LDA topic models for each question using the respective TRAINING set. During the model development, stopwords that were in the MALLET Stopword list were removed. We did not perform stemming of words and considered only single words. The LDA model requires the number of topics to be provided as an input. We conducted a preliminary analysis by providing 10 topics as input and qualitatively examining the words that constituted the topic and responses that were strongly associated with each topic. We observed that some of the topics were very similar which indicated that the optimal number of topics was fewer than 10. To determine the optimal number of topics, we used the CV_Coherence measure using the package PyLDAvis [22], as earlier studies have found CV_Coherence to be well-correlated with human judgment. We compared the CV_Coherence values of different number of topics between 5 and 10 and selected the optimal number of topics as the one with highest CV_Coherence for each question. Subsequently, LDA models were developed on the TRAINING dataset for all three questions. MALLET provides following outputs that were used for qualitative analysis:

a) A list of the top words that constitute each topic. For example, for topic T_i, the list of the top k words, $W_i = \{w_i^1, w_i^2, ..., w_i^k\}$, that constitute the topic are outputted. The value of k was set to be 20 for this study.
b) The composition of each document (open-ended responses, in our case) in terms of topics and associated weights. For example, for given topic model with n topics $\{T_1, T_2, \ldots, T_n\}$, the composition of a response R_i is represented as: $C(R_i)= p_i^1T_1 + p_i^2T_2 + p_i^3T_3+ ... + p_i^nT_n$, where p_i^j represents the relative weight associated with topic T_j and the sum of all topic weights for a document is one. Therefore, documents composed of multiple topics are expected to get assigned smaller weights for multiple topics, and documents composed of a single topic are expected to have a high weight associated for that topic.

2.2 Qualitative Analysis

The objective of qualitative analysis of the topics generated by the LDA model was to validate the understanding of underlying themes. The qualitative analysis involved the following steps:

1) First, two researchers developed initial themes for each topic from the list of top words that constituted the topic. Then, the 100 responses with the largest weights for that topic were examined to check if they corresponded to the initial theme and the themes were updated if any missing aspects were discovered. Thus, the themes were iteratively developed by sampling more instances. We selected high weight examples for theme development as they were composed mainly of a single topic of interest. To illustrate this process, one of the topics that emerged from the responses to Q3 (What would make SoP more appealing to you?) comprised the following words: *free, cheaper, cost, price, charge, expensive, print, download, version, pay, lower, certificate, online, bit, downloadable, digital, purchase, statement, pdf, copy.* By inspecting the words in context of the question asked, we can

deduce that this topic was related to the SoP cost being too expensive and a downloadable, digital copy would be a good alternative. Then, by examining strongly associated responses with this topic, such as, *"A more affordable price point. Possibly this could be done by having the option of a downloadable certificate so would save on printing, packaging, and postage,"* we could confirm that the theme we developed for the topic was appropriate but should include that a digital certificate would be considered a cheaper option.

2) The next step was to evaluate the LDA model trained on the TRAIN dataset by assessing its topic-assignment on the TEST dataset, which was not used to train the LDA model. The responses in the TEST set were manually annotated with up to three most likely topics, then checked if the top topic assigned by the LDA model was among those three. Notably, it was difficult to manually assign only one topic to responses in the TEST set, as many topics contained overlapping ideas. This is discussed further in the Results section. We also studied the relationship between the weight associated with the top topic and the level of agreement between the LDA model and human judgment.
3) We also performed qualitative analyses of responses that were composed of multiple topics according to the LDA model to further test our understanding of the topic theme. For each question, we randomly selected 100 sample cases where the LDA output had two topics with weights greater than 0.4. A researcher, who was blinded to the topic-composition assigned by the model, annotated the cases with two most prominent topics. The manual annotation was compared with the topic-composition of LDA model.
4) Sentiment analysis was performed on the responses and the sentiment-polarity of the responses associated with each topic was examined as an additional validation. We used the Natural Language Toolkit NLTK Vader sentiment intensity analyzer [23], [24], that is pre-trained on a large corpus of annotated social media text and outputs a score for Positive, Negative, and Neutral sentiments. The average sentiment score for each topic was determined by averaging the Positive, Negative and Neutral sentiment scores of responses with that as top-topic. Next, we examined whether the sentiment scores were consistent with the expected prevalent sentiment of the topic or not.

3. RESULTS

We observed the highest CV_Coherence at 6 topics for Q1 and Q2, and at 5 topics for Q3. Therefore, these were selected as the optimum number of topics and provided as input to the LDA topic model. The topics that emerged for Q1, Q2 and Q3, their themes and top-10 words, are presented in Table 1. The qualitative and sentiment analyses of topics for each question are discussed below.

3.1 Q1: Why are you interested in Statement of Participation?

The LDA model identified six topics describing interest in the SoP. Table 2 summarizes the results of qualitative and sentiment analyses for Q1. The column "%Top Topic" indicates the percentage of cases in the TRAIN and TEST datasets where that topic was the top topic. The column "%Agree-TRAIN" indicates the percentage of cases among the top 100 cases of that topic in the TRAIN dataset where the response was consistent with the theme of the topic. The column "%Agree-TEST" indicates the percentage of cases for each topic where the top topic assigned by the LDA topic model was among the three topics assigned manually. The column "Average Sentiment Score-TRAIN" indicates the average score of Positive (Pos in Table 2), Negative (Neg in Table 2) and Neutral (Neu in Table 2) sentiments as outputted by the NLTK Sentiment Intensity Analyzer for all the responses in the TRAIN dataset that had the respective topic as the top topic identified by the LDA model.

Table 2. Qualitative and Sentiment Analyses Summary: Q1

Topic	%Top Topic		%Agree		Average Sentiment Score-TRAIN		
	Train	Test	Train	Test	Pos	Neg	Neu
Q1T1	29	25	87	80	0.11	0.01	0.88
Q1T2	33	38	84	71	0.14	0.01	0.85
Q1T3	16	0	96	0	0.07	0.00	0.92
Q1T4	13	38	86	42	0.16	0.01	0.83
Q1T5	6	0	59	0	0.17	0.02	0.81
Q1T6	4	0	77	0	0.14	0.01	0.84

The agreement of the theme of the topics with human judgment in the TRAIN set was relatively good (close to 90%) for all the topics except topic Q1T5. However, we did not observe a similar level of agreement between the topic predicted by the topic model and manual annotation in the TEST set. One of the primary reasons for this effect is that the 100 responses reviewed manually in the TRAIN set had a considerably high topic-weight (>0.85) while the weights of top-topic in the TEST set were not as high (being as low as 0.28 for some cases). For the qualitative analysis of 100 responses that were mostly composed of two topics, we found that a) for 18% of the cases, the model and human judgment agreed for both topics, b) for 64% of the cases, only one of the topics assigned by the model and human agreed, and c) for the remaining 19%, neither of the two topics assigned by the model and human agreed.

Given the positive framing of Q1, the expected prevalent sentiment in learners' responses was positive or neutral, but not negative. The sentiment analysis also agrees with expectations. The responses within each topic were predominantly classified as neutral (81-92%) and positive (7-17%). It is to be noted that the NLTK sentiment analyzer, trained on annotated media corpus differing from our dataset, may produce somewhat noisy results.

Based on topic themes for Q1 as shown in Table 1, it seems that learners would be interested in obtaining the SoP if they perceive a) personal attainment value and/or a high time or effort cost for the course, for example, keeping the SoP as a memento of their hard work, b) professional utility value, such as demonstrating interest in an area to employers and universities, or c) low financial cost of the SoP and high utility or interest value of the courses, wanted to contribute back to the platform for providing great learning experiences free of charge.

3.2 Q2: If not interested in Statement of Participation, why not?

The LDA model identified six topics related to learners' disinterest in the SoP, as described in Table 1.

Table 1: List of Topics, their Themes and Top-10 Words for Q1, Q2 and Q3

Topic	Theme of the topic	Top 10 words
Q1T1	Learners wanted SoP as a proof of completing the course for personal (record of their personal achievement of finishing the course) and professional (a good addition to their resume) reasons.	record, participation, achievement, proof, cpd, completed, personal, part, add, work
Q1T2	Learners want to demonstrate their interest in a particular area for professional purposes, such as applying to universities for higher studies or demonstrating interest or skills to a potential future employer.	future, show, career, interest, proof, job, knowledge, study, university, work
Q1T3	For many learners who were working professionals, the SoP fulfilled their work-related requirement of "continuous professional development (CPD)" or training hours.	development, professional, cpd, evidence, learning, portfolio, continuing, personal, work, education
Q1T4	They wanted SoP as a reminder of the great learning experience or the time and effort they put in the course. They perceived interest or attainment value in the SoP and recognized a high time or effort cost for the course. They also wanted to show it to family and friends with pride.	time, show, learning, work, put, learn, reminder, effort, good, I've
Q1T5	Given that the courses are offered for free on the platform, learners who could easily afford to pay for the SoP wanted to support the platform so that it could continue to offer courses for free.	courses, certificate, free, pay, FutureLearn, statement, money, it's, feel, back.
Q1T6	Learners felt the SoP would be professionally useful due to various reasons, such as the course being related to their area work or coming from a reputable university.	history, university, interested, knowledge, health, teaching, work, college, education, science
Q2T1	Learners did the course out of personal interest in the subject or for leisure. They were either retired or the course was not related to their professional field.	interest, retired, personal, don't, certificate, career, participation, learning, prove, feel.
Q2T2	The price of the SoP seemed expensive to learners and they could not afford it at that time due to their financial situation.	money, purchase, buy, afford, expensive, moment, time, courses, future, cost.
Q2T3	Some learners did not need the SoP as a) they already had advanced degrees, b) they were very experienced professionally, or c) they were retired.	paper, retired, don't, certificates, knowledge, certificate, learn, learning, piece, interested.
Q2T4	Learners were not sure about the worth of SoP as it a) indicated only participation in the course and did not specify course accomplishments, learning, scores, or level of engagement, or b) was not clearly recognized by employers and universities.	certificate, participation, statement, completed, feel, complete, didnt, time, purchase, work
Q2T5	The current price of the SoP seemed high to learners due to different reasons such as their financial situation, or high currency exchange rates (if they lived in developing countries).	expensive, free, certificate, pay, cost, bit, paper, price, high, courses
Q2T6	It was difficult for international learners to buy the SoP due to high currency exchange rates and non-availability of convenient payment methods. Learners mentioned that payment through credit card or international bank transfer was not easy in their country.	card, credit, pay, payment, country, money, don't, online, bank, live.
Q3T1	Learners suggested that a) SoP should be cheaper, b) digital version of SoP should be downloadable for free, and payment should be needed for a formally verified hard copy, d) pricing should be based on the country, e) more payment methods such as *PayPal* should be supported, and f) there should be option to choose soft copy or hard copy of SoP as shipping may be difficult and costly for remote locations	free, cheaper, cost, price, charge, expensive, print, download, version, pay.
Q3T2	SoP would be more appealing if it were more relevant for their career or job, such as being recognized by employers as qualification or counting as CPD.	career, work, needed, don't, statement, job, interest, participation, retired, relevant.
Q3T3	This topic had two themes: a) the price of the SoP was too high for which some learners suggested membership model and subsidized costs for low income learners; and b) learners were not sure how to answer Q3 as some had got the SoP and some didn't want it as they did the course for recreation	courses, free, don't, appealing, money, cost, make, statement, paper, answer
Q3T4	Learners suggested that instead of showing just participation, the SoP should show detailed course achievements to properly reflect their efforts and achievements	statement, participation, certificate, completed, test, level, completion, achievement, score, tests
Q3T5	Learners would be interested in buying the SoP if it was more recognized professionally, such as course credits, recognition by employers and valid continuous professional development. Some learners suggested a more formal look of SoP with university logo.	university, qualification, recognized, credit, credits, certificate, courses, points, degree, academic

Similar to Table 2, Table 3 presents the distribution of topics, level of agreement between the topic assignment by LDA model and manual annotation, and the average sentiment scores for each topic for Q2. As shown in Table 3, the level of agreement with the human annotation in the TRAIN set is not consistently higher than the TEST set, and for some topics, it is higher for the TEST set.

Table 3. Qualitative and Sentiment Analyses Summary: Q2

Topic	**%Top Topic**		**%Agree**		**Average Sentiment Score-TRAIN**		
	Train	Test	Train	Test	Pos	Neg	Neu
Q2T1	38	58	100	79	0.11	0.06	0.83
Q2T2	25	15	80	71	0.05	0.06	0.89
Q2T3	15	12	72	100	0.08	0.07	0.85
Q2T4	12	5	65	67	0.07	0.07	0.86
Q2T5	8	6	65	100	0.08	0.06	0.86
Q2T6	3	4	93	75	0.06	0.10	0.84

Some of the possible reasons for this behavior, which is considerably different from Q1 (as shown in Table 2), may be: a) the higher number of responses for Q2 (55,000) as compared to Q1 (12,600), which may lead to samples in the TEST set being more similar to TRAIN set, and b) greater level of overlap between the topics generated for Q2 as compared to Q1. To illustrate the latter point, as shown in Table 1, there seems to be considerable amount of overlap between the themes of topics Q2T2, Q2T5, and Q2T6, with all being related to the financial cost of the SoP. This may cause the LDA model to assign either of these topics as top-topic based on the words present in the response. Additionally, these topics are highly likely to be assigned as top-3 topics during manual annotation of responses in the TEST involving cost aspect of the SoP. Therefore, it is likely to result in a higher level of agreement between manual annotation and top-topic assigned by LDA model in TEST set.

For the qualitative analysis of 100 responses that were mostly composed of two topics, we found that a) for 30% of the cases, the model and human judgment agreed for both topics, b) for 56% of the cases, only one of the topics assigned by the model and human agreed, and c) for the remaining 14%, neither of the two topics assigned by the model and human agreed. We observed higher level of agreement for top-two topics as compared to Q1.

Given the negative framing of Q2, the prevalent sentiment of responses was expected to be between neutral and negative. The sentiment scores for Q2 in Table 3 indicate that the responses were largely neutral in nature. We did not observe relatively higher score for Negative sentiment as compared to Positive sentiment (in fact, for some topics such as Q2T5, Positive had a higher average score). This differed from our expectation about the prevalent sentiment in Q2 responses.

Based on topic themes for Q2 as shown in Table 1, it seemed that learners would not opt for SoP if they perceived a) high financial or effort costs, or b) low utility or attainment value, as they did the course for leisure or did not benefit from it professionally.

3.3 Q3: What would make a SOP more appealing to you?

For Q3, the five topics that emerged from the LDA topic model are presented in Table 1. As expected, Q3 topics were similar in theme to Q2 topics, as, in Q3, learners suggested approaches to address the concerns they mentioned in Q2. Similar to Tables 2 and 3, Table 4 summarizes the distribution of topics, agreement between LDA model and manual annotation, and the average sentiment scores for Q3.

Table 4. Qualitative and Sentiment Analyses Summary: Q3

Topic	**%Top Topic**		**%Agree**		**Average Sentiment Score-TRAIN**		
	Train	Test	Train	Test	Pos	Neg	Neu
Q3T1	35	46	90	65	0.16	0.06	0.77
Q3T2	26	16	82	94	0.11	0.05	0.84
Q3T3	17	12	80	83	0.12	0.06	0.82
Q3T4	14	13	80	54	0.10	0.04	0.85
Q3T5	9	13	83	85	0.11	0.02	0.86

As shown in Table 4, we observe a high level of agreement between the LDA model and human judgment for most topics in the TRAIN set, and for all other topics except Q3T1 and Q3T4 in the TEST set. For Q3T1, the lower level of agreement in the TEST set may be due to considerable overlap in the themes of Q3T1 and Q3T3 on the cost aspect of SoP. Similarly, there is overlap in themes of topics Q3T4, Q3T5, and Q3T2 regarding the professional recognition of the SoP by employers.

For the qualitative analysis of 100 responses that were mostly composed of two topics, we found that a) for 49% of the cases, the model and human judgment agreed for both topics, b) for 44% of the cases, only one of the topics assigned by the model and human agreed, and c) for the remaining 6%, neither of the two topics assigned by the model and human agreed. We observed a higher level of agreement for top-two topics in Q3 as compared to Q1 and Q2.

Our expectation of the prevalent sentiment of Q3 responses was between neutral and positive and not as negative as Q2. The sentiment scores for Q3 responses are similar to Q1, with relatively high score for Neutral, followed by Positive, and then Negative. In summary, the learners suggested that they would be more inclined to buy the SoP if it were more affordable, recognized professionally, detailed their accomplishments and learnings; and convenient payment options were available.

4. DISCUSSION

In this study, we analyzed large number of open-ended responses using LDA topic model followed by qualitative analysis of the topics to determine and verify the topic-themes. It is important to mention the limitations associated with our study. While the topic model brought up some prominent themes from the responses, there may be other important themes that did not get highlighted because of low frequency. Therefore, the results from the topic model are not exhaustive and cannot replace detailed manual qualitative analysis that can identify such themes. It is to be noted that the topic themes were not distinct in nature and had overlapping elements with other topics, for example, in Q2, there were multiple topics on the financial cost of the SoP. During manual review process, we also noticed that learner responses often involve multiple topics and the weights assigned by the LDA model for prevalent topics may not represent the actual composition strength of the topic.

We also observed a consistent pattern for all questions that the top-topic predicted by LDA model in the TEST dataset agreed better with human annotation when the weight of the top-topic (as assigned by the LDA model) was higher. This is represented in Figure 2 as the plot between the weight of top-topic (shown as w) with agreement between LDA model and human annotation for TEST datasets of Q1, Q2, and Q3.

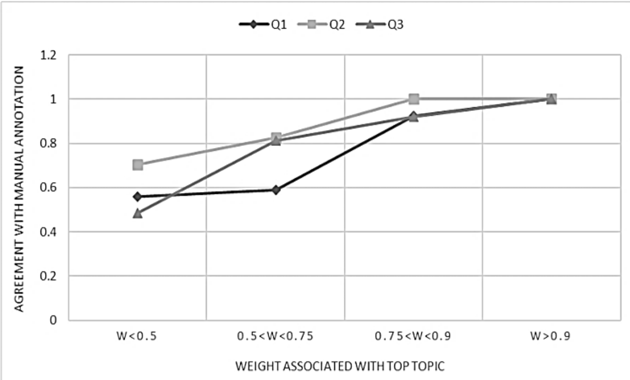

Figure 1. Weight of top-topic and level of agreement with human annotation in TEST dataset for Q1, Q2, and Q3

As shown in Figure 1, there is a relatively low level of agreement between the topic model and human judgment when the top-topic weight is less than 0.5, but picks up in the range of 0.5-0.75, and is extremely high when the weight is more than 0.75.

From the topic model analysis, there were some clear connections with aspects of value and cost in EVC theory. As expected, the expectancy dimension of motivation was not relevant for these questions. For learners for who purchased the SoP, interest, utility, and attainment values were associated with personal and career related considerations and the reputation of those offering the MOOCs, while costs were associated with task effort and time commitment. Complimentary to these findings, reasons for not purchasing the SoP were the perceived lack of value for both current and future needs, but cost focused, primarily, on the financial expense, even when high values were expressed. The suggestions for making the SoP more appealing also centered around motivational aspects of increased value and professional utility and decreasing financial or effort costs.

5. IMPLICATIONS

The implications of this study relate to the methodology of qualitative validation of topic models and learner motivations to purchase SoPs.

5.1 Methodology

Manual analysis of open-ended responses involves multiple steps such as developing a coding scheme and then coding the data, which can be challenging for large numbers of responses. Topic models provide an effective means for exploratory data analysis for a large collection of textual data but mostly require qualitative analysis for interpretability. Our results indicated that the proposed methodology for qualitative evaluation of topics generated by LDA is reliable and can be replicated for similar studies involving large-scale open-ended survey data. We also found that the topics predicted by the LDA model were more likely to agree with human judgment if the weight assigned by the LDA model was higher (>0.75). This indicates that the weight assigned by the LDA model is in line with human judgment. Still, the probabilistic nature of the LDA algorithm is such that the weights may not be perfectly representative of the composition of themes present in a response, particularly when topics are highly overlapping or consist of disparate sub-themes.

5.2 Learner Motivation

Given there is a large variation in background and learning objectives of online learners, their need for certification also varies. Research indicates that participants who pay for certification have a higher completion rate than students who choose to audit the course. Furthermore, the majority of participants report that they intend to fully participate in all aspects of the course; however, most do not fulfill this commitment. Therefore, it is important to understand what learners feel about participation certificates to improve the offering by platforms and to take advantage of the motivational benefits of certificates to increase course completion.

Based on the topics generated from learner responses, we obtained the following insights about learners' opinions of course participation certificates: a) learners were interested in buying the SoP if they valued it personally or professionally or wanted to contribute to the platform, b) learners were not interested in buying the SoP if they thought it was too expensive, lacked utility value, or were taking the course for purely recreational reasons, and c) learners believed the SoP would be more appealing if it were professionally recognized, adequately reflected effort, and cost less.

6. CONCLUSIONS

Our results showed that our multi-step approach for qualitative analysis is robust as there was high level of agreement between human judgment and topic assignment by the LDA model when the model assigned larger weight to the topic–which meant that the theme developed for the topic in the first step of qualitative analysis was appropriate. This approach for qualitative analysis of topic models would be applicable for similar studies analyzing large amounts of textual data.

This study examined how learners perceive the value of online learning certificates based on their responses to post-survey questions. It is worth mentioning that the post-course survey was taken only by learners who completed the course and not all enrollees. Future work may involve collecting feedback from all enrollees about certification in online courses that may lead to insights on their motivations for the course.

We found that one group of learners reported value in obtaining the certificate and appreciated the artifact to keep of their learning. However, another group of learners cited cost and lack of value as main reasons for not opting in for the certificate. One potential explanation may be the individual learner's socio-economic status or country location and their ability to pay for the MOOC.

MOOCs were founded as affordable learning opportunities; however, many learners indicated the certificate was priced out of their range. While obtaining a certificate may increase a learner's participation in a course and provide documentation of their achievement, it must be priced at an amount that learners world-wide can afford.

EVC theory provided a useful interpretive lens for the motivational aspects of investing in a SoP, which can be used to inform strategies for encouraging this investment and increasing course completion. Future studies could examine employer perceptions of MOOC certificates and ways of increasing the credibility of learning in a MOOC.

7. ACKNOWLEDGEMENTS

Our thanks to FutureLearn for providing us the post-course survey data. This work was made possible by the National Science Foundation (NSF) (PRIME #1544259). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

8. REFERENCES

[1] C. Wang and D. M. Blei, “Collaborative topic modeling for recommending scientific articles,” presented at the *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’11*, New York, New York, USA, 2011, p. 448. doi: 10.1145/2020408.2020480.

[2] A. Ramesh, D. Goldwasser, B. Huang, H. Daume, and L. Getoor, “Understanding MOOC Discussion Forums using Seeded LDA,” presented at the *Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications*, Stroudsburg, PA, USA, 2014, pp. 28–33. doi: 10.3115/v1/W14-1804.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *J. Mach. Learn. Res.*, vol. 3, no. Jan, pp. 993–1022, 2003. [Online]. Available: http://www.jmlr.org/papers/v3/blei03a.html.

[4] J. D. Mcauliffe and D. M. Blei, “Supervised Topic Models,” presented at the *Advances in Neural Information Processing Systems (NIPS 2007)*, 2007, pp. 121–128.

[5] D. M. Blei and D. M., “Probabilistic topic models,” *Commun. ACM*, vol. 55, no. 4, p. 77, Apr. 2012. doi: 10.1145/2133806.2133826.

[6] M. Röder, A. Both, and A. Hinneburg, “Exploring the Space of Topic Coherence Measures,” presented at the *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM ’15*, New York, New York, USA, 2015, pp. 399–408. doi: 10.1145/2684822.2685324.

[7] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno, “Evaluation methods for topic models,” presented at the *Proceedings of the 26th Annual International Conference on Machine Learning - ICML ’09*, New York, New York, USA, 2009, pp. 1–8. doi: 10.1145/1553374.1553515.

[8] J. Chang, S. Gerrish, C. Wang, J. L. Boyd-graber, and D. M. Blei, “Reading Tea Leaves: How Humans Interpret Topic Models,” presented at the *Advances in Neural Information Processing Systems 22 (NIPS 2009)*, 2009, pp. 288–296.

[9] D. O ’Callaghan, D. Greene, J. Carthy, and P. Cunningham, “An analysis of the coherence of descriptors in topic modeling,” *Expert Syst. Appl.*, vol. 42, no. 13, pp. 5645–5657, 2015. doi: 10.1016/j.eswa.2015.02.055.

[10] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, “Automatic evaluation of topic coherence,” presented at the *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, California, 2010, pp. 100–108.

[11] D. Mimno, H. M. Wallach, E. Talley, M. Leenders, and A. McCallum, “Optimizing semantic coherence in topic models,” presented at the *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Edinburgh, United Kingdom, 2011, pp. 262–272.

[12] H. Bisgin, Z. Liu, H. Fang, X. Xu, and W. Tong, “Mining FDA drug labels using an unsupervised learning technique - topic modeling,” *BMC Bioinformatics*, vol. 12, no. Suppl 10, p. S11, 2011. doi: 10.1186/1471-2105-12-S10-S11.

[13] H. C. Davis, K. Dickens, M. Leon Urrutia, M. del M. Sanchéz Vera, and S. White, “MOOCs for Universities and LearnersAn analysis of motivating factors,” presented at the *6th International Conference on Computer Supported Education*, 2014.

[14] J. Friedman, “5 Reasons to Consider Paying for a MOOC Verified Certificate | Online Colleges | US News,” 2016. [Online]. Available: https://www.usnews.com/education/online-education/articles/2016-03-04/5-reasons-to-consider-paying-for-a-mooc-verified-certificate. [Accessed: 02-Oct-2017].

[15] T. R. Liyanagunawardena, P. Parslow, and S. Williams, “Dropout: MOOC participants’perspective,” presented at the *EMOOCs 2014, the Second MOOC European Stakeholders Summit*, Lausanne, Switzerland, 2014, pp. 95–100.

[16] “FAQ- FutureLearn.” [Online]. Available: https://about.futurelearn.com/about/faq?category=certificates-and-statements. [Accessed: 28-Sep-2017].

[17] K. E. Barron and C. S. Hulleman, “Expectancy-Value-Cost Model of Motivation,” in *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, 2015, pp. 503–509.

[18] J. K. Flake, K. E. Barron, C. Hulleman, B. D. McCoach, and M. E. Welsh, “Measuring cost: The forgotten component of expectancy-value theory,” *Contemp. Educ. Psychol.*, vol. 41, pp. 232–244, Apr. 2015. doi: 10.1016/J.CEDPSYCH.2015.03.002.

[19] T. D. Reeves, A. A. Tawfik, F. Msilu, and I. Şimşek, “What’s in It for Me? Incentives, Learning, and Completion in Massive Open Online Courses,” *J. Res. Technol. Educ.*, vol. 49, no. 3–4, pp. 245–259, Oct. 2017. doi: 10.1080/15391523.2017.1358680.

[20] D. Koller, A. Ng, and Z. Chen, “Retention and Intention in Massive Open Online Courses: In Depth,” *Educause Review*, pp. 62–63, Jun-2013.

[21] A. K. McCallum, “MALLET: A Machine Learning for Language Toolkit.” [Online]. Available: http://mallet.cs.umass.edu/. [Accessed: 29-Sep-2017].

[22] “Welcome to pyLDAvis’s documentation! — pyLDAvis 2.1.1 documentation,” 2015. [Online]. Available: http://pyldavis.readthedocs.io/en/latest/. [Accessed: 21-Dec-2017].

[23] “NLTK Sentiment Analysis.” .

[24] C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text,” presented at the *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.