Introducing Context-Awareness to MOOC Systems

Mubarak Mohammad
Concordia University
Montreal, Canada
mubarak.sami@gmail.com

Alaa Alsaig
Effat University
Jeddah, Saudi Arabia
aalsaig@effatuniversity.edu.sa

Ammar Alsaig
Independent
Makkah, Saudi Arabia
saigammar@hotmail.com

ABSTRACT
Massive open online courses (MOOC) are shaping the future of education, moving studies outside class rooms, and changing the way people attain knowledge. It provide options to both universities and students. It enables students to have access to high quality courses offered by prestigious universities and top ranked professors for free or for affordable prices. At the same time, it enables universities to widen their scope and get access to worldwide prospect students. This brings in challenges to both universities and students. Universities need to address a full range of students, from different backgrounds, education levels, and specializations. Also, due to competition and variety of other MOOC service providers, universities need to establish effective techniques to persuade students to join and get engaged in its provided courses. On the other hand, students are faced with various options of universities and courses. This paper aims to tackle the challenges of 1) marketing courses for massive target audience, and 2) engaging students into the depth of the provided courses to attain effective, high quality education. We apply context-awareness techniques and principles to address these challenges. Context-Aware systems are becoming ubiquitous. These systems comprise mechanisms to acquire knowledge about the surrounding environment and adapt its behavior and service provision accordingly. Context defines the characteristics and settings of an object or event. It provides insight about the surrounding environment. In this paper, we provide a software architecture for MOOC systems which comprises context-awareness. The architecture includes two context models: one for marketing and the second for effective student engagement in the educational process. Adding context-awareness to MOOC systems helps regulate and restrict service provision to ensure effective wider reach (marketing) and better quality of education (student engagement)..

Categories and Subject Descriptors
I.1.2 [**Algorithms**]: algorithm analysis – *Computation on matrices, algebraic and non-algebraic algorithms*

General Terms
Algorithms, Measurement, Performance, Design, Experimentation,

Keywords
Context-aware; service; Architecture; MOOC.

1. INTRODUCTION
Early reports [1] show that after a year of its establishment, Coursera, which is the largest MOOC provider so far, has introduced 328 different courses from 62 universities in 17 countries. Around 3 million students registered from more than 220 countries. Six months later, the number of students doubled reaching 6 millions and 559 courses are offered from 107 universities. Courses span subjects in almost all scientific, humanities, arts, and engineering disciplines. Some courses managed to enroll more than 100,000 students. MOOC are promising to offer high quality education including best courses given by best instructors from best universities to every one on the world. MOOC platforms aim to provide real course experience, which makes it different from conventional online education, as it starts on a given date, gives real assignments and real deadlines, attempt to engage students, create study groups, include exams and provide certificates of accomplishment. Furthermore, MOOC provide additional advantages such as the ability to do exercises and refer to reference material at one's own pace at just the right time in a lecture sequence, the ability to rapidly read through transcripts of sections that are already familiar, and the ability of pause and repeat portions of lectures.

Despite the great potential and promising kick start, MOOC providers face serious challenges. The average completion rate of courses is less than 10%. Many students enroll in MOOC courses while not being sure if they want to study it or not. Only a small portion ends up completing all course workload. There are many reasons for not completing a course besides not being interested in its subject. These include: the learning experience didn’t meet expectations, not enough incentive to finish the course, forgetting about the course, being too busy to finish, loosing interest in the course, and never intending to finish the course [2]. This brings in challenges to both MOOC providers and students. It is estimated that instructors spend around 100 hours to prepare MOOC materials before the beginning of classes and 6 to 8 hours every week to follow up and engage into discussions and forums. This is a substantial investment that needs not to be misguided or abused by massive early enrolling students who never complete the course. Thus, MOOC providers need effective approaches to ensure they reach out to seriously interested students. Furthermore, MOOC providers need to employ effective techniques to engage students. Studies have shown that the deeper the engagement is the more the student retention is.

From a student perspective, high drop off rates suggest that students might not be finding what they are looking for. The huge variety of options available is accompanied with very poor searching mechanisms that provide limited options to search for available courses. For example, mooc-list.com, which is a major web site dedicated to searching for available MOOC courses, provides the following search options: category, university, length, and estimated effort. Such options are very general and makes it difficult for students to find relevant courses. Coursera

ICCASA 2014, October 15-16, Dubai, United Arab Emirates

DOI 10.4108/icst.iccasa.2014.257285

website provides the following search options: starting soon, eligible for verified certificate, language, category, university, instructor, and subject. Udacity website provides search only by category and level. Another major MOOC provider is edX, which is formed by an alliance of MIT, Harvard, Berkeley, and other top universities, provide very simple search options including only start time, category, and school. Therefore, despite the huge number of available options of courses, universities, and instructors, there is not any available effective approach that enables students to find the courses that match their interests and that enables MOOC providers to target students who would be interested in their courses. Moreover, the high drop off percentage clearly indicates the need for effective techniques to engage students and have high retention rates. This is the motivation behind the work presented in this paper. We tackle the two challenges: 1) providing effective wider reach (marketing), and 2) engaging students into the depth of the provided courses to attain effective, high quality education. Technological advancements enable MOOC providers to collect a wealth of information about every interaction with their platforms. Thus, we apply context-awareness techniques and principles to address these challenges.

The rest of this paper is organized as follows. Section II provides background information and introduces context and context-awareness. Section III introduces MOOC context. Section IV introduces our novel context-aware ranking method for MOOC to tackle the marketing problem. Section V introduces a context-awareness approach to improve student engagement and retention. Section VI provides concluding remarks.

2. BACKGROUND

In this section we introduce context. In Oxford English dictionary, context is defined as "the circumstances that form the setting of an event, statement, or idea, in terms of which it can be fully understood". According to this definition, context is necessary to fully comprehend a statement, and hence it is different from the information conveyed through the statement. As an example, the setting for a "seminar event" is a condition involving the entities: speaker, topic, time, and location. When each entity is assigned a value from the domain associated with that entity, and if the condition is met then the seminar is to be held. A condition involving n entities needs an n-tuple of values for a total evaluation. In general, many different n-tuples may satisfy a condition with n entities. So, we can regard the collection of n-tuples satisfying the condition as an n-ary relation. This is the rationale for formally defining context as a relation. We call the entities as dimensions and the values assigned to them as tag values. Note that, a tag value has a type. For example, the type of the tag values assigned to speaker is string. Therefore, context is a typed relation.

Dey et al [3] provided a different definition: "*Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application*." This definition has been adhered to by researchers in Human-Computer Interaction (HCI). The key aspect of the definition is "*relevance*", which allows the developers to choose the parameters that suit the application, including mobility.

Wan [4] has given a formal representation of context. This representation can accommodate all the above definitions, and more importantly, it is supported by relational semantics. Context is defined as a collection of ordered pairs (d, v), where d is a dimension, and v is a value from the type domain associated with dimension d. Dimensions "*Who*, *Where*, *When*, *What*, and *Why*" have been identified as primary dimensions to construct any general context. A tag set, which is typed, is associated with each dimension. As an example, along the dimension "*Where*" the tag set can be "*the set of city names or streets*", and along the dimension "*When*" the tag set can be "*the set of discreet time points*". Therefore, context is a multidimensional typed entity. An example of a context in this representation is [*Who*: Alice, *Where*: Montreal, *When*: 11: 00]. This context qualifies some events that might be experienced by Alice in Montreal at time 11. The dimensions, as suggested above, are neither selective nor exhaustive. The system designer should feel free to choose as many dimension names as are necessary. From the perspective of a design and its implementation of a specific application, the set of dimensions that are relevant for the application are to be determined by domain experts.

Example 1: *the context of a MOOC course can have five essential dimensions: Title, Category, Start Date, Instructor, and Language. Assume that Title is of type string, the tag sets (type domains) for Category, Instructor, and Language are finite sets of available categories, instructors and languages. Thus, the context c = [Title: "Social Psychology", Category: "Human Sciences", Start Date: July 14 2014, Instructor: "Scott Plous", Language: "English"] is a reference to a course context. Note that, it is possible to provide multiple values for a dimension. For example [A: 1, A: 2, B: 3, C: 4].*

Proper semantic information from the application domain must be used in choosing relevant contexts. Because contexts exist on their own, both in design and implementation, they are to be treated as first class citizens. This decision requires to model contexts as data types, with representational and operational abstractions. Using a formal representation for context is essential to enable processing of context-aware systems. A virtue of being a first class object with a concrete syntax is that context can be represented in a programming language and can be declared and created independently of other objects in the language.

3. CONTEXT-AWARE MOOC

Context-aware computing refers to a general class of mobile real-time reactive systems that continuously sense their environment and adapt their behavior accordingly. Technological advancements allow MOOC platforms to collect a wealth of information to capture the learning experience. It has been reported [2] that 230 Million data points were collected about student activities in a single course provided by edX. Also, Coursera monitors every single mouse click to analyze students' interactions. These information, if modeled and used properly, can provide important facts that can guide to improve the learning experience and face its challenges

In this section, we use context formal definition to model MOOC. This will enrich MOOC models and enable us to suggest effective matching and ranking approach to facilitate finding or promoting courses to its appropriate targets. Also, it will provide foundation for improving student engagement.

Figure 1 shows the context-aware model of MOOC. The main active entities in this model are: *university, instructor, course, student, and class*. We associate a context definition with each entity. The model includes the following components. Note that, the details of these components are beyond the scope of this paper. We focus only on the modeling part that is related to context-awareness and processing. We look at these components as black-box, self contained components without discussing details of its design or implementation.

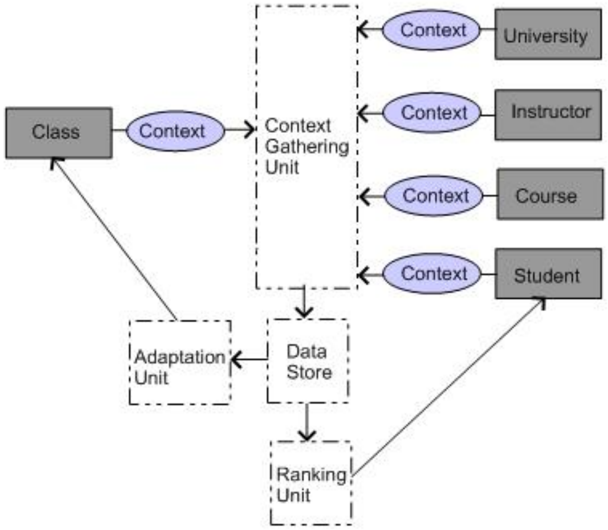

Figure 1. Context Aware Model for MOOC

- *Context Gathering Unit (CGU)*: a component that is responsible for sensing, collecting, and transmitting context information associated with each entity in the model.
- *Data Store*: a central unit that stores, manages, and facilitate querying and providing context information. The CGU continuously monitors the context information associated with each entity. Whenever any change in context information occurs, CGU sends the new updates to the Data Store. Thus, the Data Store keeps both history of context information along with direct access to the latest context information for each context.
- *Ranking Unit*: a central processing unit that is used to match and rank courses based on the context and preferences of students and the context of other system entities. Details about this unit will be provided in the next section. The objective of this unit is to solve the marketing challenge discussed in the introduction.
- *Adaptation Unite*: a role-based reactive component that monitors context updates and performs adaptations that aim to enhance student engagement. Details about this unit will be discussed in Section IV.

The remaining part of this section defines the contexts of all entities in this model. Context definitions are provided based on the formal definition of context provided in [4].

3.1 Course Context

Based on an intensive study of MOOC platforms, early analysis in [1] and [2], and knowing the essential information for effective processing of MOOC, we propose the following dimensions to define course context:

- *Title*: a string value describing the main subject of the course.
- *Category*: a predefined finite set of strings defining the main fields of study such as: Arts, Biology & Life Sciences, Business & Management, Economics & Finance, Education, Engineering, Health & Society, Humanities, Mathematics, Social Sciences, Information Technology, etc.
- *Level of Innovation*: this is an important dimension that provides a quantified measure of the level of creativity in the course subject, material, and area of study. Coursera statistics show that out of 121,000 surveyed users, 43% identified themselves as life long learners, 35% of users indicated that they enrol to learn more about the course topic, and in general only less than 10% complete courses. These numbers indicate the importance of innovation to persuade students to enrol and keep studying the subject. The tag value domain is an integer value of 1, 2, or 3 such that: 1 indicates a low level and 3 indicates a high level of innovation. Value 2 indicates medium level. Most courses by default should fall in 2nd level. Examples of highly innovative subjects offered by Coursera may include courses such as: "*Model Thinking*", "*Critical Thinking in Global Challenges*", "*Creative Problem Solving*", "*Networked Life*", and "*Social Psychology*". Each of these courses enrolled more than 100,000 students.
- *Teaching Methodology*: this is a very important dimension that describes the employed method of teaching. The tag value domain is an enumeration integer such that: 1 indicates that the teaching methodology is interactive/bidirectional, and 2 indicates that it is passive/unidirectional. It is well established that interactive courses brings in better student retention rates. An example of how interactions is implemented in MOOC is demonstrated by Coursera's platform which allows instructors to pose questions within lecture videos such that the video will pause and a pop up question appears. The student has to answer the question before proceeding to watch the rest of the video.
- *University*: string value indicating the name of the university offering the course. This is applicable only to MOOC platforms that provide university supported courses such as edX and Coursera. However, Udacity for example allows independent professionals and scientists to offer courses without affiliation to a university.
- *Instructor*: a set of string values indicating the name or names of instructors offering a course. Many MOOC courses are offered by a team of instructors, each presenting part of the course or all teaching together.
- *Quality of promotional introductory Video*: a short promo video provides the first opportunity to users to get introduced to MOOC courses and decide whether or not they will enroll. It serves as the main method of persuasion. All MOOC courses have promotional videos. The quality of the promotional video plays a key role in marketing the course and giving the first

impression about it. The tag value domain is an integer value of 1, 2, or 3 such that: 1 indicates a low level and 3 indicates a high level quality.

- *Workload*: an integer value indicating the number of hours per week the student is expected to spend doing course work. This includes the time of watching lecture videos, solving quizzes, and writing assignments. The workload is a key factor that plays dual effects. On one hand, it indicates the level of engagement with the course. On the other hand, a heavy load may cause many students to abandon the course. Thus, a balanced workload is a key factor to the success of a MOOC course.
- *Number of quizzes or exercises*: an integer number indicating the number of quizzes or exercises associated with a course.
- *Number of assignments*: an integer number indicating the number of assignments associated with a course.
- *Number of exams*: an integer number indicating the number of exams associated with a course.
- *Start time*: a date value indicating the start time of a course.
- *End time*: a date value indicating the end date of a course.
- *Accreditation*: a true/false Boolean tag value indicating the eligibility of this course to be accredited by certain universities. A true value means students can claim the credits of this course and add them to their transcript as equivalent to a course in their curriculum. Accreditation is an important factor to persuade students to take MOOC courses benefiting from its quality of education and flexible time schedule from one hand, and accounting it as part of their earned credits from the other hand.
- *Language*: a string value indicating the main language of instruction and materials used to teach a course.
- *Subtitles*: a finite set of languages for which subtitles are provided. Many MOOC platforms, such as Coursera and edX, facilitate adding subtitles to course videos.
- *Essential Prerequisites*: a finite set of courses indicating the essential prerequisites that should be taken before taking a course. For example, "*Cryptography II*", which is provided by Stanford University in Coursera, requires "*Cryptography I*" as prerequisite.
- *Optional Prerequisites*: a finite set of courses indicating some optional courses that are good to have in order to have full grasp on some sections of a given course. For example, "*Paradigms of Computer Programming*" which is provided in edX suggests a basic knowledge of programming in at least one programming language.
- *Cost*: a numeric value indicating the cost of a course. Although MOOC might give impression that courses are offered for free, it has big potential to bring wealth of revenue if price tags are carefully associated with courses. Imagine a course offered by Harvard, enrolls 100,000 students worldwide, and get priced for $10. Coursera is still pushing the idea of free education. However, in order to monetize and earn revenues to pay its expenses, it offers an option to earn a *verified certificate* for $49. The certificate provides official recognition from universities and Coursera that a student took a course and fulfilled all its requirements.
- *Past Experience*: a numeric value indicating the percentage of positive recommendations given for a specific course by previous users who took this course before. A high recommendation value indicates a successful previous experience which encourages new users to take such course once offered again.
- *Popularity*: a numeric value indicating the wide spread of remarks or comments posted about a specific course in social networks, MOOC platform, or news. The tag value domain is an integer value of 1, 2, or 3 such that: 1 indicates a low and 3 indicates a high popularity.
- *Course Format*: this is a very important dimension describing the main teaching format. Possible tag values are: short-segment videos, slides, long videos, or traditional classroom teaching. Studies [2] show that 30% of students who abandon a course state that the major reason for not continuing a course is that "*the learning experience did not meet their expectations*" or they "*lost interest in the course*". The format of teaching course materials is very important. All studies [1,2] and reviews of available information on major platforms (Coursera, edX, Udacity) suggest that using short-segment videos with each having around 6 to 7 minutes is the best approach to keep good student retention.
- *Resources Available*: a finite set of supporting materials including books, articles, documentaries, etc available for a course.
- *Schedule*: weekly scheduled time for releasing new course materials, videos, lectures, quizzes, or slides. For example, normally, Coursera releases new course materials every Monday at midnight Eastern Time.
- *Certificates*: Studies show that 11% of users who abandon courses stated that the main reason for not continuing a course is having "*not enough incentive to finish* (credits or certificate)". Certificates are important factor to encourage students to finish a course. The tag value domain is a finite set of strings.

3.2 Class Context

A class captures the complete teaching and interaction experience while giving a course on a MOOC platform. Therefore, it inherits all the context dimensions discussed above for a course. In addition, we add the following dimensions specifically for a course entity:

- *Number of students*: an integer value indicating the number of users enrolled in a course.
- *Number of Active Students*: an integer value indicating the number of enrolled students who are engaged in some activities related to a course such as: watching videos, solving quizzes, submitting exams.
- *Number of Students Completing the Course*: an integer value indicating the number of users who finish the course and accomplish a satisfactory grade to pass the course and earn a statement of accomplishment.
- *Median amount of time watching a video*: a numeric value indicating the Median time spent watching a video. It signifies the quality of teaching videos and its attractiveness to students. Shorter Median times suggest that the quality should be improved or length should be shortened.

- *Level of Interaction with Videos*: a numeric value indicating the percentage of interaction with pop-up questions embedded in course videos. High percentage indicates a high level of participation and attention.
- *Solving Quizzes*: a numeric value indicating the percentage of users who solve quizzes.
- *Solving Assignments*: a numeric value indicating the percentage of users solving assignments.
- *Taking Exams*: a numeric value indicating the percentage of users taking exams.

3.3 Student Context

We focus only on dimensions relevant to the processing of MOOC operations and enhancing its delivery. Thus, we ignore basic profile information and focus on dimensions that will affect marketing courses or enhancing engagement. We propose the following dimensions:

- *Languages*: a finite set of languages that a user masters.
- *Cost*: a numeric value indicating the cost preference of a user. It indicates the maximum amount a user is willing to pay for a course. A zero value indicates that a user is interested only in free courses.
- *Specialization*: a finite set of strings describing the background specializations and fields of study that a user accomplished.
- *Schedule*: a time value indicating the beginning of week preference for a user. In general, MOOC follows a weekly schedule that is fixed. However, users come from all over the world, from different time zones..

The contexts of *Instructor* and *University* are trivial. Therefore, we will not discuss it in this paper. The context gathering unit is responsible for collecting all contexts of all entities and synchronizing it with the data store. These information are essential for the ranking and adaptation units that we will discuss in the coming two sections.

4. RANKING CONTEXT-AWARE MOOC

The popularity of MOOC is increasing rapidly. More universities and education institutions are joining the movement. This translates to more available courses and options. For example, the number of course provided by Coursera almost doubled in the last six months. This means big number of options provided for students to choose from or even get to know about. On the other hand, the number of users is increasing at almost the same rate. For example, the number of users enrolled in Coursera doubled from around 3 millions to 6 millions in the past 6 months. These numbers are based on the officially announced statistics on their website. This means that universities and course providers have huge variety of audience with different backgrounds and specialities. Thus, reaching out to the proper audience becomes more challenging. A symptom of this problem is the high drop out of students who enrol in courses and then find out that they are not interested very soon in the first week. This is because there is no proper matching approach that suggests the best suitable courses for students. Often, universities advertise all their courses to all students who enrolled in any of its courses, which means random targeting that sometimes becomes annoying to MOOC users and may make them unsubscribe or ignore marketing advertisements that might include a course that they might be interested in. Another issue contributing to this problem is the lack of agreed upon standards or conventions for naming or categorizing courses. This makes searching for courses based on simply matching queries about course titles or categories, which is the main approach so far, a non-useful and unproductive solution. There is not a single approach for handing this problem so far neither in research nor in MOOC platforms

In this section, we propose a novel approach for ranking MOOC courses based on the context model that we introduced in the previous section.

We will refer to the process of searching for relevant courses by student or searching for suitable target students by MOOC platform as the *discovery process*. Context is multidimensional. Therefore, the discovery process should support multiple features. The domain types of tag values are heterogeneous (integers, numeric, strings). In the following subsections, we will introduce few important concepts:

- *Query*: the starting point of any discovery process is defining a query which indicates relevant fields and required values for each field. Every field value corresponds to a context dimension. We define query using the same approach used to define context. This will make it easy to use it for the discovery process. The following example shows a simple query:

  Query = [Category: "*Social Sciences*", Cost: 0, Workload: 5 hours]

- *Weights*: during discovery process, the user may have different preferences and levels of importance associated with each value. We use weights to model the level of importance given to each field in the query. These weights will make difference and help differentiate high volumes of interesting options. We use numerical values from 1 to 5 to indicate the selected weight for a specific query dimension such than 1 indicates low importance and 5 high importance. For example, if one is searching for any free course then the following weight vector corresponds to the previously defined query:

  Weight = [1,5,3]

- *Matching*: matching is the process of finding possible answers to a query. It puts a query against all possible options and selects the exact options that correspond to the required values. For example, all the following courses match the stated query:
  - [*Category*: "Social Sciences", *Title*: "Moralities of Everyday Life", *Popularity*: "Low", Cost: 0, *Workload*: 5],
  - [*Category*: "Social Sciences", *Title*: "Social Psychology", *Popularity*: "High", *Cost*: 0, *Workload*: 4],
  - [*Category*: "Information Technology", *Title*: "Cryptography I", *Popularity*: "Low", Cost: 0, *Workload*: 3]

  Simple matching will result only in the first option because it satisfies all requested dimensions. However, the student will miss out an opportunity to attend a more popular course that has even less workload as demonstrated in the second option.

- *Ranking*: the process of classifying and ordering search results based on its approximate closeness to a defined query is called ranking. The difference between matching and ranking is that matching filters out all other options which does not exactly match a predefined dimension value. However, ranking considers all available options but orders them according to their approximate closeness to a given query while respecting defined weights of importance for each dimension such that a higher rank indicates a very interesting option and a lower match indicates a less important option. Hence, it does not exclude options and at the same time differentiate options based on their relevance and importance.

- *MOOC Context* is the composition of all the context dimensions defined in the previous section.

There exists a high volume of literature about ranking. It has been used in many domains. However, there is no ranking approach that is suitable for MOOC platform. In this paper, we introduce a new multi-featured context-dependent similarity measure formula that is designed to fit MOOC. The formula is vector based.

Before introducing the ranking formula we will classify MOOC context dimensions into four categories:

- *General dimensions*: such as title, category, Teaching methodology, university, instructor, accreditation, language, course format. These information must be matched with exact values. Not matching any value will decrease the ranking of an item.
- *Quality dimensions*: such as quality of promotional video, popularity, past experience, certificate, number of students completing the course, Median time watching videos, number of students solving quizzes, number of students enrolled in a course, and number of students taking a course. The higher the value of these dimensions the better.
- *Overhead dimensions*: such as workload, quizzes, exams, assignments, prerequisites, and costs. Lower values of these dimensions are better.
- *Time-based dimensions*: such as start date and schedule. A period of time must be defined for these dimensions. If a tag value occurs in this period then it is regarded as a match.

In the following explanation and examples, we will focus only on the scenario in which a student is discovering the available courses. The other scenario of MOOC platform promoting a course to a student follows the same procedures. Thus we skip it in this paper due to size limits. Our ranking approach comprises the following steps:

1. Preparing query and weight vectors,
2. Computing discovery measure for each dimension,
3. Forming discovery measure matrix,
4. Computing the product of the discovery measure matrix and the weight vector,
5. The result is a row matrix the gives the ranks of all alternatives.

The discovery measure formula receives the query vector, weight vector, and MOOC context as input. Then for each dimension *Qi* in the query vector, it compares its value to an available alternative context dimension *Di* of an available course, and it produces one of the following possible results:

- 1:
    - If the dimension *Di* is *General* and its value equals the value of *Qi*
    - *If the dimension Di is Quality and its value equals the value of Qi*
    - *If the dimension Di is Overhead and its value equals the value of Qi*
    - *If the dimension Di is Time-based and its value is in the specified period.*
- $\frac{Di}{Qi}$:
    - *If the dimension Di is Quality and its value is less than the value of Qi. For example, if we have two past experience dimensions: 60 and 70 and if the query dimension value is Qi = 75 then the resulted measures are* $\frac{60}{75}$ *and* $\frac{70}{75}$*. Thus, the resulted measures are: 0.8 and 0.93. We observe that the highest quality gets the best measure.*
- $1 + \frac{Di}{Max(Di)}$
    - *If the dimension Di is Quality and its value is bigger than Qi, where Max(Di) is the maximum available alternative dimension value for this specific dimension. For example, if there are three possible levels of interaction with video dimensions: 70, 80, and 90. Then, Max = 90. In this case the calculated discovery measures will be:*$1 + \frac{90}{90}$, $1 + \frac{80}{90}$, $1 + \frac{70}{90}$*, the resulted measures are: 2, 1.89, 1.78. We observe that the highest quality gets the best measure.*
- $1 + \frac{Min(Di)}{Di}$
    - *If the dimension Di is Overhead and its value is less than Qi, where Min(Di) is the minimum available alternative dimension value for this specific dimension. If Di is 0 then we replace the fraction with 1. For example, if there are three possible Cost dimensions: 10, 20, and 30. Then, Min(Cost) = 10. In this case the calculated discovery measures will be:*$1 + \frac{10}{10}$, $1 + \frac{10}{20}$, $1 + \frac{10}{30}$*, the resulted measures are: 2, 1.5, 1.3. We observe that the lowest workload gets the best measure.*
- $\frac{Qi}{Di}$:
    - *If the dimension Di is Overhead and its value is bigger than the value of Qi. For example, if we have two Overhead dimensions: 4 and 6 and if the query dimension value is Qi = 2 then the resulted measures are* $\frac{2}{4}$ *and* $\frac{2}{6}$*. Thus the resulted measures are: 0.5 and 0.3. We observe that the lowest overhead gets the best measure.*

- *0*
    - *If the dimension Di is General and its value does not equal the value of Qi.*

After computing all discovery measures for all dimensions, we formulate a *discovery measure matrix*. For simplicity we will run a basic example just to show how to formulate this matrix. Our approach is capable of measuring complete ranking for all MOOC context. Note that the power of our approach is not demonstrated by such simple example because including more dimensions and options results in better view of the power of the ranking approach. However, due to page size limits we show a query of only three dimensions.

Assume we have the following course contexts:

- C_1 = [*Category*: "Social Sciences", *Title*: "Social Psychology", *Popularity*: 5, *Workload*: 4],
- C_2 = [*Category*: "Social Sciences", *Title*: "Models of Life Experiences", *Popularity*: 2, *Workload*: 8],
- C_3 = [*Category*: "Computer Science", *Title*: "Introduction to C# Programming", *Popularity*: 3, *Workload*: 3].

Assume we have the following Query: [*Category*: "Social Sciences", *Popularity*: 4, *Workload*: 3], and we have the following weight vector: [*Category*: 5, *Popularity*: 1, *Workload*: 4]. The first step is to calculate the discovery measure and formulate the matrix. The following matrix shows the calculated measures:

$$\begin{bmatrix} 1 & 1 & 0 \\ 2 & \frac{2}{4} & \frac{3}{4} \\ \frac{3}{4} & \frac{3}{8} & 1 \end{bmatrix}$$

Such that the columns represent the discovery measures for each course and the rows represent the values for each dimension *Di* for all courses. For example, the calculated measures for C1 ap[re [Category = 1, Popularity = 2, Workload = $\frac{3}{4}$]. These calculations are based on the formulas stated above.

The next step is to calculate the matrix product of the Weight

$$\text{Final Rank} = [\,5,1,4] \times \begin{bmatrix} 1 & 1 & 0 \\ 2 & 0.5 & 0.75 \\ 0.75 & 0.375 & 1 \end{bmatrix}$$

= [5 + 2 + 3, 5 + 0.5 + 1.5, 0 + 0.75 + 4] = [10, 7, 4.75]

Thus, the ranking values for C1, C2, C3 are 10, 7, 4.75 respectively. The result indicates that C1 is more suitable and closer to what the student is looking for than C2 and both are closer than C3. More complex examples with more options can be calculated using the same approach. Hence, this approach can be used for marketing courses based on the knowledge of current MOOC context and queries or preferences of students and MOOC providers.

5. PERFORMANCE EVALUATION

The algorithm introduced in this paper considers high performance. Given the high number of courses MOOC platforms can contain, it is crucially important for the ranking algorithm employed for MOOC platforms to obtain the results in a timely manner. In this paper, we examine our novel ranking algorithm considering two cases*; scale out* case, and *scale up* case. *Scale out* case examines the performance of the ranking process when the number of dimensions increases, where the *scale up* case examines the algorithm when the number of courses increases. These are the only factors that can have effects on the performance of the algorithm.

In *scale out* case, we have performed 10 executions of the ranking algorithm. For each execution, we fixed the number of available courses to 10,000 courses, and increased the number of dimensions by 10 in every execution, starting the first execution by 10 dimensions. Meaning, the first execution was with 10 dimensions, the second was with 20 dimension, and so on until the 10^{th} execution, which we ended up with 100 dimensions and 10,000 courses. Figure.2 shows that the algorithm exhibits a linear growth.

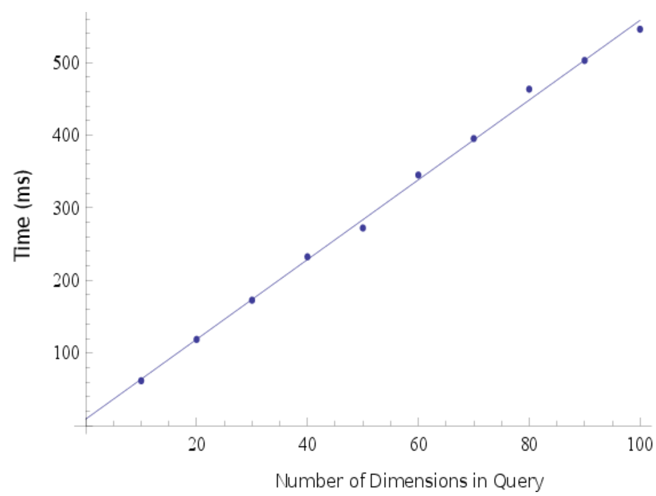

Figure 2. performance of the algorithm when number of dimensions increases

In *scale up* case, we have followed the same procedure, but we have fixed the number of dimensions to 6 and made the number of courses changeable. We have also performed 10 different executions in this case. The evaluation started with 10,000 courses and 6 dimensions. With the addition of 10,000 to the number of courses in each execution, our final execution contained 100,000 courses and 6 dimensions. Figure.3 shows the algorithm performance when the number of courses increases.

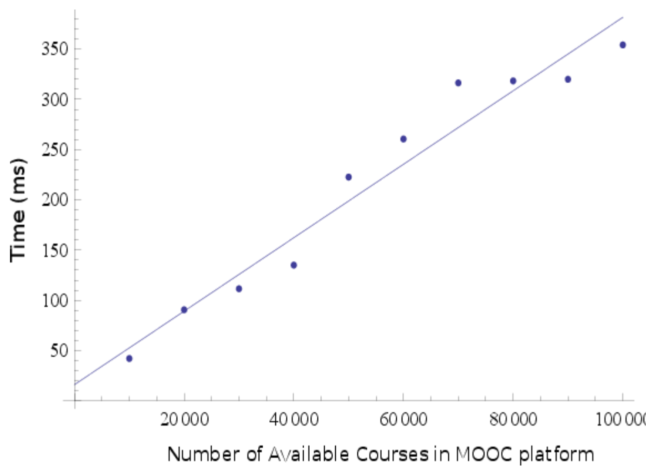

Figure 3. performance of the algorithm when number of courses increases

The difference of numbers in the two cases is because of the following reasons:

1. In *Scale out* dimensions are what define the available courses. It should be entered, viewed and weighted by the potential student as a query. Thus, the number of dimensions should be handful. Otherwise, defining a query might be a difficult task. Nevertheless, for an accurate evaluation, we have considered a number of dimensions up to 100, which is much more than the norm

2. In *Scale up* the norm of the number of courses is to be high. As introduced earlier in this paper, number of courses can reach millions. Thus, we have considered high numbers in scale up case

In conclusion, both scenarios, i.e. scale up and scale out, exhibit linear behaviour. This proves that our ranking algorithm is able to rank courses in a timely fashion. Thus, it is applicable to MOOC platforms

6. ADAPTATIONS TO IMPROVE ENGAGEMENT

The knowledge of context at any point of time can help increase the quality of the learning experience and service provision. In order to succeed and survive the intense competition between different MOOC providers, MOOC platforms should be adaptive. Adaptation means changing strategies, service delivery, and quality of service continuously. Here changes mean improvements and suitable adjustments to enhance the learning experience. Adaptation depends primarily on gathering and processing context information. Our proposed Context Gathering Unit should continuously monitor and collect dimensions of all engaged entities in the model. After that, rule-based reasoning engines should be configured to analyze the overall context of MOOC. MOOC platform should include a flexible rule-based engine that uses context to evaluate situations and execute rules that are befitting to a situation. For example, if the Median time of watching videos decrease then this is an indicator that perhaps video-segments are long or have boring materials. Another scenario is: if only some videos have lower Median, whereas others have higher Median than it indicates that those videos with higher median have interesting materials that should be focused on or effective engagement techniques that need to be reused. If the number of students enrolled in the course who solve quizzes or assignments decrease then this might be an indicator of heavy workload or less interest in continuing the course. At this point, providing some incentives might be useful to increase student retention. Many other scenarios and rules can be defined in the same manner.

Employing adaptation is not a trivial task and it might be overwhelming. The collected statistics drawn from context information might be misleading sometimes. For example, if the number of students engaged in a course decreases this might be an indicator that many of them enrolled by mistake or just for exploring the topic in general. This means that it is normal that the median time of watching decreases. In this case, the minority of students who actually enrolled because they wanted to complete the course might be enjoying it and the quality of delivery for them might be acceptable. Thus, in such scenarios perhaps no adaptation is required and the course better continues as it is. However, this requires a method of discriminating those students who are really interested by the course and those who are just passing by. The ranking algorithm that we provided in the previous section can play an important role in this process. It will allow the MOOC platform to select only the students who have better disposition to enjoy and finish the course by ranking all students and selecting only the higher ranked portion as target for selective adaptations. The lower ranked portion should not be ignored but rather different adaptations are employed for them. Hence, by employing context gathering, ranking, and rule-based reactions, it is possible to perform for the same course different adaptations that might result in better learning experience for all participant each according to their level of interest. This approach can help instructors to focus only on interested students and not waste their energy assuming that they are serving 100,000 students whereas in fact they are actually serving only 5000.

7. CONCLUSION

This paper introduces a novel context-aware approach for improving the quality of delivering MOOC services. At the core of this approach, we proposed a model and context definitions for MOOC entities. The context comprises comprehensive dimensions that are important to provide MOOC services. Then, we provided a novel ranking algorithm that uses MOOC context and defined queries to discover and rank best options for students and MOOC providers. By employing this approach, students can find best suited courses, and MOOC providers can effectively target interesting students. Finally, we provided high level guidelines of how context and ranking can be used to enhance the engagement of students. We suggest that MOOC platforms continuously adapt to the rapid changes happening during the learning and teaching experience. A smart rule-based engine can provide valuable suggestion of improvements based on the current status of the system.

MOOC are currently at the infancy stage with great potential and big challenges. We believe that employing context-awareness techniques can take MOOC to next levels of higher quality of delivery.

8. REFERENCES

[1] M. Mitchell Waldrop, "Online learning: campus 2.0, massive open online courses transform higher education and providing fodder for scientific research," Nature Magazine, Vol 495 Issue 7440, pp. 160-163, March 2013.

[2] Geoffrey A. Fowler, "An early report card on massive open online courses," The Wall Street Journal, October 8 2013.

[3] Anind K Dey, Gregory D Abowd, and Daniel Salber, "A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications," Human-computer interaction, 16(2):97–166, 2001.

[4] Kaiyu Wan, "Lucx: Lucid enriched with context," PhD thesis, Concordia University, 2006.