MOOC-O-Bot: Using Cognitive Technologies to Extend Knowledge Support in MOOCs

Ameesha Mittal[a, b], Lorenzo Vigentini[a], Mentari Djatmiko[a], Gangadhara Prusty[c], Yashvardhan Sharma[b], and Mark. E. King[a]

[a]*Office of the Pro Vice Chancellor Education*
UNSW Sydney
Sydney, Australia
{l.vigentini, mentari.djatmiko, mark.king}@unsw.edu.au

[b]*Department of Computer Science and Information Systems*
BITS Pilani
Pilani, India
{ff2014107, yashg}@pilani.bits-pilani.ac.in

[c]*School of Mechanical and Manufacturing Engineering*
UNSW Sydney
Sydney, Australia
g.prustyg@unsw.edu.au

***Abstract*—There is an abundance of intelligent tutors and expert systems across all domains. Some are driven by "cognitive technologies" and conversational agents, which have become common in providing an accessible interface to a range of devices. They are often used to provision initial support for businesses to guide users of their websites through frequently asked questions and help material. However, there are only few examples applied to the educational context, and Massive Open Online Courses (MOOCs) in particular. This paper presents work in progress aiming to bring cognitive technologies to use in the running of MOOCs and supporting both instructors and students to answer content-related questions. The overall architecture of the system is presented, together with the modeling of the knowledge base in three different MOOCs from different disciplines. The approach is promising in offering a tool capable of providing teacher support, as well as direct student support, in dealing with the large amount of questions posted in discussion forums and in answering common questions regarding both the running of the course as well as content.**

Keywords—MOOC, cognitive technologies, conversational agents, online learning support, question and answers

I. Introduction

Two streams of work converge in the work presented in this paper: the exciting rise and proliferation of MOOCs [1][2] bringing a wealth of interaction data; and the long tradition of the Artificial Intelligence in Education (AIED), Interactive Tutoring Systems (ITS), Educational Data Mining (EDM), Learning Analytics (LA), Learning Sciences communities and the more recent acceleration of cognitive technologies. MOOCs are online courses offered free or at a very low cost by commercial platform providers (of which Coursera, EdX, FutureLearn and Udacity are the most well-known), often developed in partnership with Universities or established industry experts with a variety of educational designs reaching millions of learners each year [2]. Considering intelligent systems, Baker [3] captured the state of automatic tutoring: "One of the initial visions for intelligent tutoring systems was a vision of systems that were as perceptive as a human teacher (see discussion in Self 1990) and as thoughtful as an expert tutor (see discussion in Shute 1990), using some of the same pedagogical and tutorial strategies as used by expert human tutors (Merrill et al. 1992; Lepper et al. 1993; McArthur et al. 1990)." His thesis focused on the potential of educational data mining as a supporting tool driving *human* decision-making and supporting online learning, with a focus on intelligence "amplification" rather than artificial intelligence *per se*. The need for such augmentation is essential in fully online courses, and MOOCs in particular, in which the aspiration of a fully connected learning network is often trumped by technical limitations of the existing platforms [4][5] and the individual motivations of learners taking these courses varies greatly.

A. Problem Statement

Within the backdrop of these intersecting fields there are two important issues to tackle. Teachers in MOOCs have to make sense of all activity, engagement and interactions and are often unable to make the most of the "teacher presence". On the other hand, students are often facing feeling of confusion and loneliness as they are not able to get the *right* answers when they need them [6]. Thus, MOOCs have opened up a new pool of opportunities for using cognitive technologies to improve the educational experience. We aim to develop a system relying on conversational technology to be deployed in MOOCs that can serve as a teaching support for instructor or as automated teaching assistant capable of handling students doubts related to any MOOC domain content.

II. Background

A MOOC provides a way to bring together the people interested in gaining knowledge (the *students*) and people or *experts* who seek to provide this knowledge [7]. Among several tools on the MOOC platforms, discussion forums are an important source of student engagement and a primary source of interaction among students as well as between students and tutors. Students being active on the discussion forums can act as a good indicator of their engagement [7]. Consequently, these discussions forums produce a large amount of data available for researchers to work with.

A concern often raised about MOOCs is low completion rates: although several thousand enroll for courses, very few complete such courses [1][2][8]. There are several reasons identified for this problem, one of them being a feeling of isolation and lack of interactivity experienced by the students. Normally students try to clear their doubts via posting on discussion thread or emailing the tutors. Issues that are quickly clarified in a classroom can remain outstanding for several

978-1-5386-6522-0/18/$31.00 ©2018 IEEE 4-7 December 2018, Wollongong, NSW, Australia
2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)

hours or even days in online courses and MOOCs. Looking at reasons for non-completion in MOOCs, Khalil and Ebner [8] indicated that 35% of students were not satisfied by the presence of *proper* instructor interaction. They also felt that fellow students were not as helpful as in traditional classroom setting. The key problem seems that tutors find it difficult to scale up support, monitoring hundreds of posts and responding in a timely manner.

In the field of AI, the hope to use natural language to enable machines to interact with humans has been around for several decades. Yet, it is only in the past five years that the proliferation of consumer tools relying on speech recognition and "chatbots" or "digital assistants" like Siri (Apple), Cortana (Microsoft), Alexa (Amazon), Google Assistant (Google), Watson (IBM) have become common. In simple terms, a conversational agent or a chatbot is a service enabled by rules and some form of artificial intelligence that can engage in an interactive conversation with humans. These technologies are at a point in which they mimic conversation rather than understand it [9] and the systems are far from perfect, with some equating conversational agents to bad PAs [10]. Skimming through the past few decades, from the creation of ELIZA (1960s), the first conversational tool using a pattern matching and substitution methodology [11] provided a tool which could pass the Turing test for a few utterances. This prompted visionaries to experiment the possible applications in education with Computer Assisted Instruction. In 1970 Carbonel presented the first intelligent tutoring system, SCHOLAR [12]. In the 1980s, researchers shifted to intelligent tutoring systems (ITS) [13], focusing on cognitive models. Anderson described the separation of procedural from declarative knowledge in the ACT model [14]. In the 1990s, scholars in AI started to discuss embodied cognition as a *sine qua non* true intelligence. Over time, the focus shifted from pure pattern matching in utterances, to rule-based system with a student knowledge model [15] or automatic student models with cognitive tutors [16].

The applications of these conversational agents are extensive, and within the educational domain some notable examples are presented in the next sections. At the time of writing, there are primarily two possible approaches for building a conversational agent: rule-based and corpus-based. Corpus based approach can be further classified as information retrieval-based or generation-based [17].

A. Practical Implementations

Although there are some fundamental differences between traditional Intelligence Tutoring systems, which heavily rely on the modelling of the domain knowledge and the students, there are several examples of implementation of cognitive technologies in the educational domain with varying degrees of sophistication. Carbonell's SCHOLAR [12] was one of the first using AI in education, envisaging the idea of a system with a database consisting of some form of knowledge related to any course or subject along with generic knowledge of language and some principles of tutor instruction. AutoTutor presents a chain of challenging problems and assists students in finding the correct answer by drawing upon his/her knowledge and adaptively correcting the problems with the appropriate responses [18][19]. It supports Natural Language conversation and uses latent semantic analysis for the analysis of student's responses. AutoTutor improves upon student's knowledge by comparing the students' responses to text paragraphs in its database (representing learning expectations or goals) and creating a multidimensional domain model. Comparison is done using latent semantic analysis. It then uses the outcome of this analysis to decide upon a suitable "dialogue moves" like an assertion or a hint etc. It works as a standalone application that act as tutor on the whole course and drives the flow of conversation. A recent meta-analysis [20], showed an impressive effectiveness as an instructional tool, with learning gains over non-interactive learning materials on a variety of math and science domains. Another system, Oscar [21] focused on adapting the mode of instruction by matching student's learning style to provide better tutoring. Specifically in MOOCs, the use of conversational agents to make online learning experience better is exemplified in [22], which presents an Artificial Intelligence Markup Language (AIML) using a third party software and web speech API. *Jill Watson* refers to a family of bots created to act as virtual teaching assistants in MOOCs is closely related to our ultimate goal. Built in 2016, the first version of Jill Watson used IBM Watson APIs and provided a memory of questions-answers pairs from the discussion forums of previous runs of the course organized into multiple categories [23]. Further, the tool was deployed and tested only on one course and it is to be determined if the technology used is transferable to other courses. Outside of education, Lokman and Zain [24] presented a query flow based chatbot system, ViDi, for detecting diabetes in a patient and suggesting suitable diet. In addition, Ferguson [25] showcased a conversational system to solve routing problems in a simple transportation domain and several other such researches present dialogue system for multiple use cases.

B. Rule Based Approaches

In the rule-based approach, a conversational agent answers questions based on some pre-defined rules. The most commonly used language to develop conversational agents is AIML, a language based on XML; both ELIZA [11] and ALICE [26] used this architecture. The advantage of using this approach is its accuracy: as matches of a predefined pattern find corresponding excerpts, this ensures a correct reply. Moreover, the replies produced by this approach are likely to be free of any grammatical or spelling errors.

The major limitation of this approach is the difficulty of scaling. Because the entire interaction is engineered, a lot of time and effort is required to implement the system; this makes it very costly to design conversations. Another limitation is that the system can only reply to the inputs that match with some pattern on which it is trained. It does not allow for inconsistencies in conversational patterns and the sophistication of non-verbal interaction..

C. Corpus Based Approaches

These use Natural Language Processing (NLP) techniques to mine responses from provided corpora of data. The corpora can either be from chat or online discussion forum data, or news or articles, like Wikipedia. Corpus based approaches can be broadly categorized in two categories: information retrieval-based systems and supervised machine learning-based systems using sequence transduction (generation based methods [17]). Information retrieval based systems try to respond to the input text by extracting the most relevant statement from the corpus of available data. These are more scalable than rule-based systems since all the responses are retrieved from existing documents, which guarantees fluency and contextual validity. However, they lack the accuracy provided by rule-based systems. Generation-based systems try to create a response by transducing from the users' prior turns to the system's turn. These systems are generally modeled using sequence to sequence (seq2seq) models [27] and are based on an encoder-decoder framework. The input text is encoded into a vector representation and fed to the decoder. The decoder then generates response based on how the neural networks have been trained. Since such systems 'generate' the answers by taking word by word input, they seem more intelligent than systems based on other approaches. However, they are also more prone to spelling and grammatical errors, require large amount of data to train, and are extremely CPU intensive.

D. *Combining Multiple Approaches for Chatbot Architecture*

In the literature, some argue that combining different architectures creates more efficient systems. For instance, Song et al. [28] proposes an ensemble of retrieval based and generative chatbot systems by feeding the output of retrieval-based system in the generative system along with the query to provide the generative system with more information. Another system, called *AimeChat* combines the retrieval based and generation based model and uses an attentive Seq2Seq re-rank model to optimize the combined results [29]. Although these examples come outside the educational domain, they suggest interesting avenues that will provide the major contribution of this paper.

III. PROPOSED METHODOLOGY

The novel hybrid model presented here focuses primarily on the way knowledge extraction and generation is achieved. The proposed architecture combines the rule-based and information retrieval-based approaches for dialogue systems. Figure 1 shows the overall architecture of the proposed approach, with attention to the handling of content-related queries, rather than any query in general. This core feature is the first stage of development for a conversational agent in MOOCs capable of supporting students' queries, but also augmenting teachers' ability to tackle queries at scale. First, we detect queries from the past comments of the MOOC to create a corpus of past questions using a query detector. This query detector is trained using the NPS chatroom dataset [30] and labeled comments from the course *Through Engineer's Eyes*. Second, we construct a knowledge-base from the course content (video transcripts, course material and past comments of the MOOC). The two datasets are then used to generate suitable responses for user query using four main components: 1) a keyword extraction component, 2) a question classification component, 3) an information retrieval (IR) component and 4) a pattern-matching component. The keyword extraction component extracts the keywords from the query using the dataset of past queries. The keywords extracted are then passed to the query classification component along with the keywords extracted from the knowledge base. This component identifies the questions as being content-related or generic (generic questions include technical difficulties with the platform, queries about the running of the course such as quizzes or assignments, and general conversational questions such as "How are you?" and "What's your name?"). The content related questions are sent to the IR component which searches for relevant responses from the knowledge base. The generic questions are handled by the pattern matching component. In this paper, focus only on the part that deals with content related queries.

A. *Query Detection*

To parse the queries from past comments we developed a model with two classifiers. For each comment in the input corpus, the model classifies it as being a question or not. The first part of our model is a Naive Bayes classifier trained on the nps_chat corpus [30]. This component classifies individual sentences of the comment in one of the 15 dialogue act types: Accept, Bye, Clarify, Continuer, Emotion, Emphasis, Greet, No Answer, Other, Reject, Statement, System, Wh-Question, Yes Answer, Yes/No Question. Thus the comment is now represented as a 15 valued vector where each value represents the number of sentences belonging to a particular dialogue type. This vector is then fed into a logistic regression classifier trained on labeled comments from the course *Through Engineers Eyes*. The logistic regression classifies the binary option of whether the comment is a question or not.

B. *Knowledge Base*

Our knowledge base consists of chunks of data formed from three data sources: 1) the transcripts of the video lectures, 2) the uploaded course material and 3) the comments from previous runs of the course. A *chunk* is a piece of consistent information about any topic that can serve as a response to a question related to that topic. Chunks are formed as follows:

- From the video transcripts, three consecutive sentences form a chunk. Thus, each sentence lies in at least one chunk (if it is the last or the first sentence) and at most in three chunks.
- From the course material, each paragraph is considered as a chunk of data.
- From the past comments, each comment is first preprocessed to remove any salutations, greetings, names etc. and then each preprocessed comment is treated as a chunk.

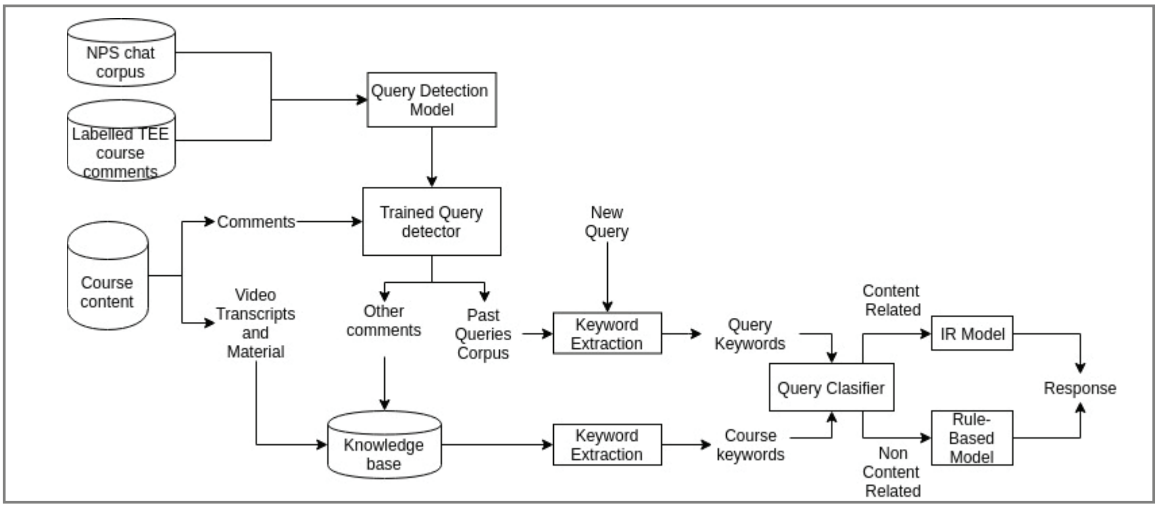

Fig. 1. Complete Architecture of the proposed approach to build the knowledge base for questions-answers, essential to drive a MOOC conversational agent.

C. Keyword Extraction

We have used Term Frequency - Inverse Document Frequency (TF-IDF) based approach [31] to extract keywords from the chunks in the corpus as well as input questions. TF-IDF is one of the most popular keyword extraction algorithms when a document corpus is available [32]. The technique determines how important a word is in a document based on the frequency of that word in the document as well as in the whole corpus. Generally, the more frequently a term occurs in a document the more important it is in that document. However, the more the word occurs in the whole corpus the less important it is to any individual document. The weight of any term i in a document j in a corpus of D documents is mathematically given by Equation (1).

$$weight_{i,j} = freq_{i,j} \times \log(D/doc\_freq_i) \quad (1)$$

We have used L2 normalization to normalize the TF-IDF scores returned by the above formula for any document. In order to extract keywords from a query, we treat each query as an individual document and all the past queries as the document corpus. For extracting keywords from the chunks in the knowledge base each chunk is treated as a document and the whole knowledge base, as the corpus. In either case, if a term has a normalized TF-IDF score > 0.1, it is treated as a keyword. The extracted keywords are used for determining the type of query and then extracting relevant answer for the query from the knowledge base in case the query is content related.

D. Query Classification

Based on the keywords extracted for a question and those extracted from the knowledge-base we label the question as content-related or non-content related. We then define a set of "system words", which includes terms such as Chrome, operating system, link and browser, to identify platform related issues like any technical difficulties in submitting quizzes, opening any link etc. We find the intersection of the question keywords with the keywords generated from the course content for that particular module as well as with the system words. If the intersection with the course words is more than that with the system words, it is treated as a content related query.

E. Information Retrieval

Our information retrieval component pulls relevant responses for any question from the knowledge base using a four-stages approach: preparation, text matching, validation and learning. This is based on term co-occurrences [33] between the question and individual chunks of data in the knowledge base. In the training phase, the system learns from user feedback to improve upon its performance.

1) Preparation

Since our model is based on term co-occurrences, this needs at least two keywords from the question in order to extract relevant responses from the knowledge base. In case the question posed by the user is not clear enough (i.e. contains no keywords), the system asks for more information from the user. It then obtains the keywords for the added information using the keyword extraction model described in Section III.C. The keywords obtained are used by the following stage along with the keywords extracted for all the data chunks in the knowledge base.

2) Text Matching

Our approach treats the questions and the knowledge base chunks as bag of words. The extraction begins with finding common keyword pairs between the question and each data chunk. For the relevance score calculation, only these keyword pairs are used. We do not consider single keywords or triples, quadruples, etc. Each common pair P with keywords k_1 and k_2 between a chunk c and query q is given a pair weight calculated using (2)

$$w((k_1, k_2), \mathrm{c}) = icf(k_1, k_2)\times(ks(k_1, q, c)+ks(k_2, q, c)) \quad (2)$$

where icf stands for *inverse chunk frequency* of the pair and ks stands for keywords score for individual keywords in the pair. The score for any keyword k in a common pair is calculated using normalized TF-IDF scores (represented by S) of that keyword in the query and in the chunk along with an importance factor θ (which is discussed in the next section) as follows:

$$ks(k, q, c) = \theta_{k,c} \times S_{k,q} \times S_{k,c} \quad (3)$$

The query-chunk relevance score is the summation of pair weights of all common keyword pairs (k_1 and k_2) between the query q and the chunk c. It is calculated by (4).

$$score(c) = \sum w((k_1, k_2), c) \quad (4)$$

Using these formulas, the system calculates the relevance scores of all the chunks belonging to that module of the course for which the query was asked. Ten data chunks with the highest relevance score are then forwarded for validation.

3) Validation

Since this system tries to retrieve an answer from a limited knowledge base, it is possible that none of candidate responses chosen are relevant enough to serve as the answer to the user question. Thus we need to compare the relevance scores of the candidate responses to a threshold value before displaying them to the user. Currently this threshold value is zero i.e. as long as the data chunk has some relevance score, after sorting the top 10 most relevant chunks, these are displayed to the user. However, in future work, we aim to determine a more appropriate threshold value which includes both the absolute value of questions asked to the system and user-entered relevance.

4) Learning

This component aims to improve the quality of responses returned to a user based on users feedback. When the suitable candidates are displayed as possible answers to a user question, the system asks the user to choose the best response, if any. This increases the relevance of a specific response with respect to that query. To do so we define an *importance factor*, represented by θ in (3). By default, the value of θ is set to 1. If a user chooses response *r* as the best response, the θ value of all the common keywords between the question *q* and response *r* is increased by a factor of the normalized TF-IDF score of that keyword in the question. The same can be represented mathematically as follows for every common keyword *k* between *q* and *r*:

$$\theta'_{k,r} = \theta_{k,r}(1+S_{k,q}) \quad (5)$$

This ensures that the relevance score of the chosen chunk would be higher next time if a similar question is asked. In case none of the returned responses is good enough, future implementations will enable the user to enter a valid response, which would be added to the knowledge base with the theta value of the common keywords updated as per (5). This will enable teachers to expand the knowledge base.

IV. Results

In this section, we present the results of the different components of the system using three datasets.

The first test case is an engineering MOOC (Through Engineers' Eyes) which is used to describe the process. Then, a similar approach is extended to two additional MOOCs: *Personalised Medicine* and *International Franchise Law*, which demonstrate the portability and relevance of the approach.

A. Query Detection

As explained in section III.A, our query detection model uses two classifiers. A Naive Bayes classifier creates a generic approach to identify questions. The Classifier was trained using 10,000 posts from the NPS Internet Chatroom Conversations dataset [30]. This classifier identify dialogue types of individual sentences in a comment. The second classifier (logistic regression classifier) aims to provide weights to different dialogue types. It was trained using a manually labeled dataset from the Through Engineers' Eyes MOOC. The total labeled dataset consisted of 2342 comments from the first run of the course and 812 from the second run with 276 and 136 queries respectively. We used the data from the second run of the course to train the model and that from the first run of the course to test the model. In the process to identify the best approach, we experimented with several classification models and compared the results. These models include:

- Model 1: Naive Bayes model trained only on the labeled course comments.
- Model 2: Decision Tree trained only on the labeled course comments.
- Model 3: Naive Bayes model trained only on NPS chatroom dataset using *wh-question* and *yn-question* dialogue act types to identify questions.
- Model 4: Naive Bayes model trained on NPS chatroom dataset using *clarify*, *wh-question* and *yn-question* dialogue act types to identify questions.
- Model 5: The used query identification model described in section III-A.

We measured the accuracy, precision and recall for each of these classifiers (Table I). Since the test data contained a majority of comments not classifiable as questions, the recall and precision scores are more informative than accuracy [34]. As shown, model 5 (chosen) achieved better performance than all the other models across the three scores. Models trained only on the labeled dataset from past comments were able to achieve good accuracy, however they have zero recall and precision. This happened because the test data used contained proportionally many more non-questions comments than comments classified as questions. Even with this inherent bias, the models produced good accuracy scores in matching relevant candidates for the limited number of questions detected in the training phase.

TABLE I. Performance of Different Query Identification Models with Performance Metrics from the Test Dataset

Model	Accuracy	Precision	Recall
Model 1	91.30%	0.00%	0.00%
Model 2	91.36%	0.00%	0.00%
Model 3	93.84%	45.00%	73.69%
Model 4	68.00%	65.62%	16.15%
Model 5	94.30%	74.27%	76.78%

TABLE II. EXAMPLES OF QUERIES AND EXTRACTED KEYWORDS USING THE PROPOSED APPROACH

Course	Text	Keyphrases Extracted
Through Engineer's Eyes	Does this experiment have any similarities to real life situations? If so which one? And what sort of jobs would lead to conducting this experiment?	['real life situations', 'similarity', 'sort of jobs', 'experiment']
Through Engineer's Eyes	maybe I'm looking at a different question. What step number was it?	['different question', 'step number']
Through Engineer's Eyes	Creeping is caused by the outward force(centrifugal) exerted on the rotor blades of the gas turbine engines at high temp. They can cause catastrophic situations if not inspected. On the other hand, can creeping be also associated with train track lines in a linear force?	['catastrophic situations', 'creeping', 'gas turbine engines at high temp', 'linear force', 'outward force', 'rotor blades', 'tain track lines', 'other hand']
Through Engineer's Eyes	Please correct me if wrong: So, the deflection curve shows two behaviours, depending on the load? I mean, as a soft one when the load is small, or as a stiff one; when the load is bigger. What do you mean by Undamped? What is the difference between a linear and nonlinear springs?	['load', 'deflection curve', 'nonlinear springs', 'linear', 'undamped', 'behaviour', 'difference']
Personalized Medicine	Then it's true that sometimes cancer disease passes directly to the second generation. And if you have fybrochistic mastocytis? What are the probabilities of developing breast cancer?	['cancer disease', 'fybrochistic mastocytis', 'probability', 'second generation', 'breast cancer']
Personalized Medicine	Did I miss a section where coding and pairs were introduced?	['coding', 'pair', 'section']
International Franchise Law	Good advice in how trademarks are protected in different countries. The one thing that i would like to emphasize on is using of environmental friendly resources. I would like to add into human labor, using the locals in the community around might work to the advantage of the franchise as they may also advertise it to the people around. Also it creates a good raptor with the community around	['community', 'environmental friendly resources', 'good advice', 'good raptor', 'thing that i', 'human labor', 'different countries', 'advantage', 'local', 'trademark', 'people', 'franchise']
International Franchise Law	To conduct due diligence As a master franchisee I would make sure if there is any disclosure documents and other collateral documents in the franchisor's own language. Is there any be practical and legal reasons why using the local language? As a master franchisee I will make sure that I do not engaged in any activities that could violate the Foreign Corrupt Practices Act or local anti bribery statutes.	['master franchisee', 'foreign corrupt practices act', 'legal reasons', 'local anti bribery statutes', 'local language', 'other collateral documents', 'own language', 'disclosure documents', 'activity', 'due diligence', 'franchisor']

TABLE III. EXAMPLES OF QUERIES FROM DIFFERENT DATASETS AND IDENTIFIED CLASS USING THE PROPOSED APPROACH

Course	Query	Class
Through Engineers Eyes	Windows 10 Professional (x64) Version 1607 (build 14393.321), Firefox 49.0.1 - Anything else you need to know? Not sure what you mean with etc.	Non-Content Related
Through Engineers Eyes	Talking of which, where is everybody? The number of comments in this and the previous week are very few indeed.	Non-Content Related
Through Engineers Eyes	Referring to the previous quiz, if friction is independant of area of contact, how is the rope around a bollard experiment explained?	Content Related
Through Engineers Eyes	what's the main difference between non-linear and linear spring ? kindly provide practical example.	Content Related
Personalised Medicine	Um. This exercise won't launch on my laptop?	Non-Content Related
Personalised Medicine	Is there a possibility that, despite testing positive for BRCA1, the breast cancer in the example scenario could have developed unrelated to the faulty gene? If so, is there a risk of treating for the wrong type?	Content Related
International Franchise Law	Is it possible to download those diagrams from video in order to examine them carefully?	Non-Content Related
International Franchise Law	Fast-food franchisees like McDonalds, KFC and Burger King appear to be favourites. Can anyone suggest a reason why this might be the case?	Content Related

B. Query Classification

The query classification module determines if the question is content related or non-content related. Due to the lack of labeled data we couldn't quantify the performance of this module, however answers were evaluated by tutors and Table III shows examples of the questions for the three courses.

C. Information Retrieval

Question-answer pairs from previous instance of three courses from different disciplines were used to test the information retrieval component. We selected the question-answer pairs where the answers were provided by the instructor of the course. We generated the responses for those questions and checked if any of the top 10 candidate responses matched the response of the instructor. Since our knowledge base contains comments from the past runs of the courses, ideally the system should have returned the exact same comments of the tutors. Unfortunately, this validation technique did not produce very good results. The likelihood of finding the same response given by the tutor was 29.7% for the course *Through Engineers' Eyes*, 25.6% for the course *Personalised Medicine* and 30.1% for the course *International Franchise Law*. While the system did not return the exact same response, the candidate answers returned for many questions were still relevant. Table IV presents some exemplars of relevant, but non-matching answers generated. This shows that the only *true* test of the accuracy of the system will require experts rating the answers returned; this will be done in the next stage of development.

TABLE IV. EXAMPLES OF RESPONSES RETURNED FOR THE QUESTIONS WHERERESPONSE DID NOT MATCH TUTOR RESPONSE

Through Engineer's Eyes	
Question	What is undamped motion?
Relevant responses returned by the system.	Damping would mean something akin to friction. I presume that makes any motion come to a stop after the initial force has been removed. a sort of decay in the amplitude of movement. A undamped spring would oscillate for ever. as stated in the video.
Question	Actually. I didn't get the exact difference between static and dynamic loads. Can anyone explain it to me?
Relevant responses returned by the system.	1) Not absolutely sure. but I think the difference is about whether acceleration is involved. When you add the load and gently lower the pan. statics explains what happens. If you apply the load and let the pan fall. it accelerates under gravity and then bounces about a bit. It will stretch further than in the slow application of force and then recoil to a shorter extension. Dynamics will explain that. Is that right. do you think? 2) Ok. So i think I got the difference static is not moving. Dynamic is moving. The spring bounces to twice the length that it would stretch if we let it down gently. 3) the difference between static and dynamic conditions is the presence of acceleration. 4) Understand the difference from the physical point of view Static load = (The sum of all forces is zero (0)), means that the object is not moving. Dynamic load = (The sum of all forces is equal to mass times acceleration (m * a)). in this case (a = gravity) because it is accelerating the object means that the object has displacement, velocity and acceleration. I hope I have helped.
Personalized Medicine	
Question	So can a person have cancer without having the type of gene configuration which could pass on to their children?
Relevant responses returned by the system.	1) I learnt that if a person has cancer. it doesn't necessarily mean they have the type of gene which can be passed to future generations. The ethical scenarios were thought- provoking. A scenario where a person has a high chance of developing cancer (due to a hereditary gene) but has not actually got the disease, and is olfered risk-reducing surgery (and they have dependant children) seems a particularly diliicult dilemma. I would like to know more about personaising drugs to reduce cholesterol and blood pressure. but I guess as cardiovascular disease and strokes are related to lifestyle and environment. it would be dilficult to know what levels would be considered 'nonnal'. ie. reducing risk of the event happening. 2) There are many genes that cause dilierent types of cancer. BRCAgenes are just one of many examples. BRCA genes are located in within chromosomes 13 and 17 and can be inherited both maternally and paternally. There can be other genes which are mitochondrial and hence only maternally inherited that can still predispose to cancer.
International Franchise Law	
Question	Can I protect/monopolise my brandfbusiness idea worldwide and be in total control of that niche market and my 'know-how'?
Relevant responses returned by the system.	1) The very important aspect. in my opinion. of any business venture is the degree of control the owner retains over the total operation of the business. I would therefore prefer the appointment of a master franchisee in the country I want to grow in. The masterfranchisee's agreement would ensure that l. as a franchisor. retain as much power and control as legally possible. 2) I believe if you are to expand into a country you need to look at the local businesses and how they are ran. and then compare how you run yours and how you would like your franchise to be run. If you do not agree with the workers pay. protection use of ethical products then you either have to enter a niche market in this area and change that, with careful monitoring to make sure that that franchise remains based on your values or find another place to expand to.

V. CONCLUSIONS

After an extensive review of the past literature we proposed a novel architecture for a conversational agent which can serve in MOOCs to handle user questions related to any course. We developed a generic approach to identify questions asked by students from a corpus of comments from previous MOOC offerings, which uses keyword extraction to identify the nature of the question and then applies an information retrieval model on content related questions. This model used keyword co-occurrences to extract potential answers to the question from a knowledge-base including all available sources as course material, video transcripts and past comments. Furthermore, the system is capable of improving its results based on user feedback. We also discussed some experiments and results demonstrating the efficiency of the question detection and information retrieval models in the system. The system was tested on a corpus of 2342 comments with 276 queries and it was able to achieve an accuracy of 94.3%, precision of 74.3% and recall of 76.8%. The validation of the IR model relied on instructor answered question-answer pairs for three courses from different disciplines. We proved that even though the system was only able to achieve an average accuracy of 28.9% with this validation method (exact same responses as the tutors provided), it was still able to generate alternative, relevant responses for the questions from the corpus available. Future work will focus on 1) extending validation approaches for the IR model, 2) a rule-based model to handle non-content related questions and 3) a thorough testing of the effectiveness of whole system with the help of experts rating the responses returned by the system for different questions. The system will be deployed in a real MOOC to obtain user feedback on the quality of answers and improve the labelled dataset for training. Further, the tool could benefit from speech recognition and more interactive UI, enabling the *bot* to communicate more like humans and better cater people with accessibility problems. The conversational agent could interact with users by combining text, voice and a graphical user interface.

Even if this work is only a start in the direction of bringing AI in MOOCs we hope that the early success of our work can encourage others in further explore the range of applications of hybrid architecture conversational agents and benefit both tutors and students in MOOCs and online learning.

REFERENCES

[1] K. Jordan, “Initial trends in enrolment and completion of massive open online courses,” *The Int. Review of Research in Open and Distributed Learning*, vol. 15, no. 1, pp. 134–60, Jan. 2014.

[2] D. Shah, “By the numbers: MOOCS in 2016 — class central,” Class Central’s MOOC Report, 25-Dec-2016. [Online]. Available: https://www.class-central.com/report/mooc-stats-2016/.

[3] R. S. Baker, “Stupid tutoring systems, intelligent humans,” *Int. J. of Artificial Intell. Educ.*, vol. 26, no. 2, pp. 600–614, Feb. 2016.

[4] S. Bayne and J. Ross, "The pedagogy of the massive open online course (MOOC): the UK view". York: Higher Education Academy. Available: https://www.heacademy.ac.uk/system/files/hea_edinburgh_mooc_web_240314_1.pdf, 2014.

[5] C. Penstein Rosé *et al.*, “Challenges and opportunities of dual-layer MOOCs: Reflections from an edX deployment study,” in *Proc. of the 11th Int. Conf. on Comput. Supported Collaborative Learning*, 2015, vol. 2.

[6] F. J. García-Peñalvo, V. F. Hermo, Á. F. Blanco, and M. Sein-Echaluce, “Applied educational innovation MOOC: Learners’ experience and valorization of strengths and weaknesses,” in *Proc. of the Second Int. Conf. on Technological Ecosystems for Enhancing Multiculturality*, New York, NY, USA, 2014, pp. 139–145.

[7] T. R. Liyanagunawardena, A. A. Adams, and S. A. Williams, “MOOCs: A systematic study of the published literature 2008-2012,” *The Int. Review of Research in Open and Distributed Learning*, vol. 14, no. 3, pp. 202–227, Jul. 2013.

[8] H. Khalil and M. Ebner, “MOOCs completion rates and possible methods to improve retention - a literature review,” in *Proc. of EdMedia + Innovate Learning 2014*, 2014, pp. 1305–1313.

[9] R. P. Schumaker, M. Ginsburg, H. Chen, and Y. Liu, “An evaluation of the chat and knowledge delivery components of a low-level dialog system: The AZ-ALICE experiment,” *Decision Support Systems*, vol. 42, no. 4, pp. 2236–2246, Jan. 2007.

[10] E. Luger and A. Sellen, “‘Like having a really bad PA’: The gulf between user expectation and experience of conversational agents,” in *Proc. of the 2016 CHI Conf. on Human Factors in Computing Syst.*, New York, NY, USA, 2016, pp. 5286–5297.

[11] J. Weizenbaum, “ELIZA—a computer program for the study of natural language communication between man and machine,” *Commun. of the ACM*, vol. 9, no. 1, pp. 36–45, Jan. 1966.

[12] J. R. Carbonell, “AI in CAI: An artificial-intelligence approach to computer-assisted instruction,” *IEEE Man-Mach. Syst.*, vol. 11, no. 4, pp. 190–202, Dec. 1970.

[13] J. R. Anderson, C. F. Boyle, and B. J. Reiser, “Intelligent tutoring systems,” *Science*, vol. 228, no. 4698, pp. 456–462, 1985.

[14] J. R. Anderson, “ACT: A simple theory of complex cognition,” *Amer. Psychologist*, vol. 51, no. 4, pp. 355–365, Apr. 1996.

[15] J. Kay. Gauthier, C. Frasson, and K. VanLehn, “Stereotypes, student models and scrutability,” in *Intell. Tutoring Syst*, 2000, pp. 19–30.

[16] M. Eagle *et al.*, “Exploring learner model differences between students,” in *Artificial Intell. in Educ.*, 2017, pp. 494–497.

[17] Z. Yan *et al.*, “DocChat: An information retrieval approach for chatbot engines using unstructured documents,” in *Proc. of the 54th Annual Meeting of the Assoc. for Computational Linguistics (Volume 1: Long Papers)*, Berlin, Germany, 2016, pp. 516–525.

[18] A. C. Graesser *et al.*, “AutoTutor: A tutor with dialogue in natural language,” *Behavior Research Methods, Instruments, & Comput.*, vol. 36, no. 2, pp. 180–192, May 2004.

[19] A. C. Graesser, P. Chipman, B. C. Haynes, and A. Olney, “AutoTutor: an intelligent tutoring system with mixed-initiative dialogue,” *IEEE Trans. Edu.*, vol. 48, no. 4, pp. 612–618, Nov. 2005.

[20] B. D. Nye, A. C. Graesser, and X. Hu, “AutoTutor and family: A review of 17 years of natural language tutoring,” *Int. J. of Artificial Intell. Educ.*, vol. 24, no. 4, pp. 427–469, Dec. 2014.

[21] A. M. Latham, K. A. Crockett, D. A. McLean, B. Edmonds, and K. O’Shea, “Oscar: An intelligent conversational agent tutor to estimate learning styles,” in *Int. Conf. on Fuzzy Systems*, 2010, pp. 1–8.

[22] S. L. Lim and O. S. Goh, “Intelligent conversational bot for massive online open courses (MOOCs),” arXiv:1601.07065 [cs], Jan. 2016.

[23] A. K. Goel and L. Polepeddi, “Jill Watson: A virtual teaching assistant for online education,” Georgia Inst. of Technology, Tech. Rep., 2016.

[24] A. S. Lokman and J. M. Zain, “An architectural design of virtual dietitian (ViDi) for diabetic patients,” in *2009 2nd IEEE Int. Conf. on Comput. Sci. and Inform. Technology*, 2009, pp. 408–411.

[25] G. Ferguson, J. Allen, and B. Miller, “TRAINS-95: Towards a mixed-initiative planning assistant,” in Proc. of the Third Int. Conf. on Artificial Intell. Planning Syst., pp. 70–77, 1996.

[26] B. A. Shawar and E. Atwell, “ALICE chatbot: Trials and outputs,” *Computación y Sistemas*, vol. 19, no. 4, pp. 625–632, Dec. 2015.

[27] L. Shang, Z. Lu, and H. Li, “Neural responding machine for short-text conversation,” arXiv:1503.02364 [cs], Mar. 2015.

[28] Y. Song, R. Yan, X. Li, D. Zhao, and M. Zhang, “Two are better than one: An ensemble of retrieval- and generation-based dialog systems,” arXiv:1610.07149 [cs], Oct. 2016.

[29] M. Qiu *et al.*, “AliMe chat: A sequence to sequence and rerank based chatbot engine,” in *Proc. of the 55th Annu. Meeting of the Assoc. for Computational Linguistics (Volume 2: Short Papers)*, Vancouver, Canada, 2017, pp. 498–503.

[30] E. N. Forsythand and C. H. Martell, “Lexical and discourse analysis of online chat dialog,” in *Int. Conf. on Semantic Computing (ICSC 2007)*, 2007, pp. 19–26.

[31] S. Robertson, “Understanding inverse document frequency: on theoretical arguments for IDF,” *J. of Documentation*, vol. 60, no. 5, pp. 503–520, Oct. 2004.

[32] B. Lott, “Survey of keyword extraction techniques,” UNM Education, vol. 50, 2012.

[33] E. Sneiders, “Text retrieval by term co-occurrences in a query-based vector space,” in *Proc. of COLING 2016, the 26th Int. Conf. on Computational Linguistics: Tech. Papers*, Osaka, Japan, 2016, pp. 2356–2365.

[34] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets,” *PLOS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015.