FAQtor : Automatic FAQ generation using online forums[*]

Ankita Bihani
Stanford University
Stanford, USA
ankitab@stanford.edu

Jeff Ullman
Stanford University
Stanford, USA
ullman@stanford.edu

Andreas Paepcke
Stanford University
Stanford, USA
paepcke@cs.stanford.edu

ABSTRACT

Forum content is an accumulation of significant human effort. It is thus sensible to extract prolonged benefit from this investment. We propose a system to automatically generate course specific frequently asked question lists (FAQs) from multi-year forum datasets. The system creates two separate FAQs: a Student FAQ, and an Instructor FAQ. The Student FAQ provides help for queries asked in the past. The Instructor FAQ provides an overview of the forum activity in previous course offerings. We present two models for Student FAQ generation using logistic regression and random forest classifiers. The logistic regression classifier reaches an accuracy of 68%. The OOB estimate of error rate for the random forest classifier is 31.48%. The predictors are easily obtained features from forum facilities, such as upvotes, unique views, and unique collaborations. For ground truth, we used expert judgment by current teaching assistants for the same course.

Keywords

Online Discussion forums, MOOCs, residential courses, FAQ, logistic regression, LDA, topic modeling, random forest.

1. INTRODUCTION

As institutions of higher education seek to expand their geographic coverage, online discussion forums have been the most obvious tool at hand. Massively open online classes (MOOCs) deployed them immediately along with the medium of instructional video. These discussion forums empower instructors and students to engage one another in ways that promote critical thinking, collaborative problem solving and knowledge construction [13] [16]. In the case of geographically distributed learning populations, no time of physical collocation for discussion is available. In such settings, discussion forums have been a primary means for information flow, outside the unidirectional video stream [1].

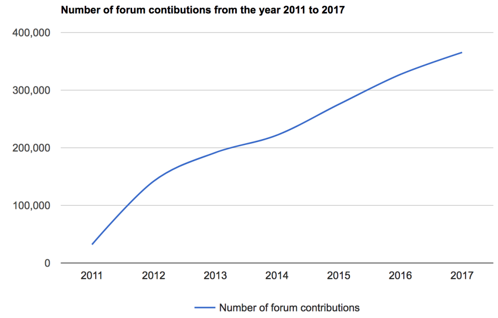


Figure 1: Growth in the number of Piazza forum contributions for courses at our University over the years

However, several residential courses also make heavy use of the discussion forums. The need for students to ask questions, voice concerns, or to point out errors in course material are as salient in residential settings, as they are in less traditional situations, such as distance learning [3]. Figure 1 shows the rapid increase in contributions to just one of the several available online forum tools, over the past years, in a large private university.

Several years' worth of a course's forum archive hold treasures for a number of stakeholders. The answers to many relevant questions might be buried in those archives, and would save time for future students and teaching staff alike, if they were made available, in some way. Instructors could learn from those archives where students tend to falter, and thus tailor their lectures accordingly. In fact, forum content is an accumulation of very significant human effort. It is therefore sensible to extract long-term benefit from this investment.

Unfortunately, this potential is not tapped from today's use of forum facilities. Questions are often re-raised, because their earlier answers are unavailable. This limitation motivates the construction of course specific frequently asked questions lists (FAQs).

However, it can take an unrealistic amount of time and effort to manually comb forum archives for question–answer pairs

[*](Does NOT produce the permission block, copyright information nor page numbering). For use with ACM_PROC_ARTICLE-SP.CLS. Supported by ACM.

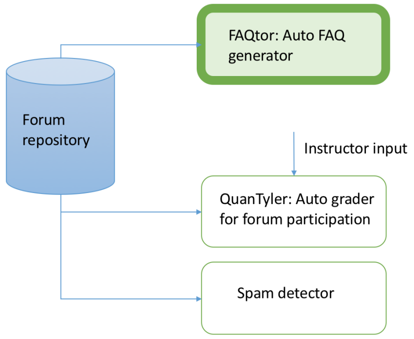

FAQtor: Auto FAQ generator
Forum repository
Instructor input
QuanTyler: Auto grader for forum participation
Spam detector

Figure 2: Block diagram of our framework around forum facilities

worth including in a specially curated list. Thus, an automated construction and maintenance of frequently asked question lists from live and archived forum posts is a more promising proposition.

In this paper, we present the details of FAQtor, our automatic FAQ generation system. FAQtor is one of the components in a larger coherent system under construction by the authors. That system will boost the value of online class forum facilities. As shown in Figure 2 the system also includes support for automatically assigning academic credit for forum participation [4], and a spam detection component.

Broadly, our contribution in this paper is: (i) to show the prediction performance of a logistic regression model in selecting forum posts for FAQs. We also show the training accuracy of a random forest model. (ii) to show the minimal random forest that reaches a stable error rate, and (iii) to show the relative importance of the readily available predictors that can be used for automatic FAQ selection.

2. RELATED WORK

Over the years, online discussion forums have become a primary focus of educational research [15]. Forums provide a fertile ground for collaborative learning and engagement in online courses. Many researchers have strongly endorsed the importance of discussions in collaborative learning [2][18]. This interaction, captured in the transcripts of the threaded discussion forums, is an object of active research. In this paper, we propose a system to automatically generate course specific FAQs using the forums archives.

FAQs satisfy three broad goals. Firstly, they provide the end-users with an easy access to browse the key information to provide them instant help or solution to their query. Secondly, they aid the party responsible for answering the queries by saving them the trouble of answering the same query multiple times. Thirdly, they help reduce the influx of repetitive questions on the forum which can soon make the forum unwieldy and difficult to navigate through.

FAQ Finder [8], Auto-FAQ [21] and Automated FAQ Answering [20] by Sneiders are some of the representative FAQ retrieval works. In [23], the authors proposed a knowledge share platform structured as a FAQ, which is constantly enriched by the newly generated interactions in the community. When a learner raises a question through the user interface designed for the system, a list of candidate answers are returned from the FAQ knowledge base. If none of the responses are satisfactory, the learner can post the question to the *community*, and wait for a response. Finally, this new question–answer pair is added to the FAQ knowledge base.

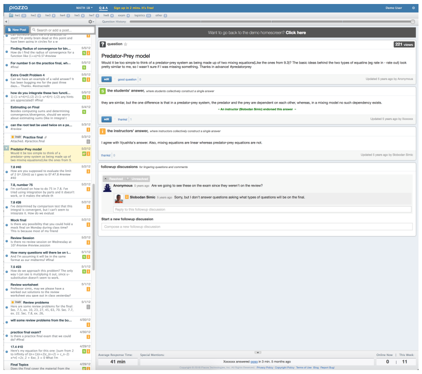
More recent work by A.Moreo et al.[14] proposes a framework to provide high quality FAQ retrieval systems. The two retrieval algorithms used are: Minimal differentiator expressions algorithm and TF-IDF. What sets this work apart from its previous related works is the fact that FAQ managers also have the usage reports at their disposal to monitor the FAQ performance and tailor the FAQ accordingly.

These studies assume the pre-existence of a solid FAQ knowledge base as a starting point, and focus on aspects like retrieval from FAQ, maintenance, and improvement of the FAQ. In fact, with a few exceptions, most of the previous work [8], [21], [20], [12],[23],[14],[11] in the FAQ world, focuses on FAQ retrieval from an existing FAQ knowledge base, rather than creating and identifying questions for the FAQ knowledge base itself.

Hu et al.in [9] propose a semi-automatic method to identify similar questions posted in Open Source Project forums in order to assist forum managers in constructing the FAQ. Questions in the forum are clustered into several groups. Each clustered group represents a set of similar or related questions. These clustered groups, are then presented to the forum managers, by suitable visualizations methods to help them construct the FAQ. The primary limitation of this approach is that the system only *aids* the forum managers and reduces their effort in selecting the FAQ. The manual selection and maintenance of FAQ is still the responsibility of the forum managers.

To the best of our knowledge, the most closely related work to our paper is [19]. In [19], the authors used hierarchical agglomerative clustering to identify the FAQ. A distance metric was defined to harness similarities based on bag of words and word embeddings. The authors extracted questions asked by students from Khan Academy[10], and a FAQ was extracted for each topic. Our work is different from [19] in two broad aspects:

- In [19], the authors focus on inferring similarity between questions and grouping similar questions into clusters. Thus, this approach is expected to work well only in settings where similar questions are very often re-raised by students (within the same course offering). However, in forums like Piazza, used at our University, we observed that the average number of near-duplicate questions within a single course offering were roughly under 1%. Thus, we cannot rely on similarity between questions to select posts relevant for the FAQ. Our FAQ creation approach relies on easily obtained features from forum facilities, such as upvotes, unique views and collaborations.
- In [19], the authors extract only questions and cluster



start? I'm pretty brain dead at this point and have been going in circles for a w

Finding Radius of convergence for bin... 5/5/12
How do I find the radius of convergence for a function like (1+x)^0.5? #review

For number 5 on the practice final, wh... 5/5/12
#final

Extra Credit Problem 4 5/5/12
Can we have an example of a valid answer? It has been bugging me for the past three days... Thanks. #extracredit

how do you integrate these two functi... 5/5/12
1) (1-x^4)^(1/2) 2) (1-x^4)^(-1/2) any hints are appreciated!! #final

Estimating on Final 5/4/12
Besides computing sums and determining convergence/divergence, should we worry about estimating sums (like in integral t

can the root test be used twice on a pa... 5/3/12
#review

Instr Practice final 5/3/12
Attached. #practice.final

Predator-Prey model 5/2/12
Would it be too simple to think of a predator-prey system as being made up of two mixing equations(Like the ones from 9.

7.8 #40 5/2/12
How are you supposed to evaluate the limit of 2 (t^.5)ln(t) as t goes to 0? #7.8 #review #40

7.8, number 75 5/2/12
I'm confused on how to do 75 in 7.8. I've tried using integration by parts and it doesn't work, or it makes the whole th

7.8 #26 5/2/12
I've determined by comparison test that this integral is convergent, but I can't seem to integrate it. How do we evaluat

Mock final 5/2/12
Is there any possibility that you could hold a mock final on Monday during class time? This is because most of my friend

Review Session 5/2/12
Is there no review session on Wednesday at 10? #review #review.session

How many questions will there be on t... 5/1/12
And I'm assuming it will be in the same format as our midterms? #final

7.5 #23 5/1/12
How do we approach this problem? The only way I can see is mutiplying it out, since u-substitution doesn't seem to work.

Review worksheet 5/1/12
Professor simic, may we please have a worked out solutions to the review worksheet you gave out in class yesterday?

Instr Review problems 5/1/12
Here are some review problems for the final: Sec. 7.5, ex. 10, 23, 27, 45, 63, 70. Sec. 7.7, ex. 22. Sec. 7.8, ex. 26.

will some review problems from the bo... 4/30/12

practice final exam? 4/30/12
Is there a practice final exam that we could do? #final

17.4 #10 4/29/12
Here's my equation for this one: [sum from 2 to infinity of ((n+1)(n+2)c_(n+2) + c_n-2) x^n] +2c_2 + 6xc_3 = 0 What I'm

Final Topics 4/29/12
Does the Final cover the material from the

Predator-Prey model

Would it be too simple to think of a predator-prey system as being made up of two mixing equations(Like the ones from 9.3)? The basic ideas behind the two types of equatins (eg rate in - rate out) look pretty similar to me, so I wasn't sure if I was missing something. Thanks in advance! #predatorprey

Updated 5 years ago by Anonymous

the students' answer, *where students collectively construct a single answer*

they are similar, but the one difference is that in a predator-prey system, the predator and the prey are dependent on each other, whereas, in a mixing model no such dependency exists.

~ An instructor (Slobodan Simic) endorsed this answer ~

Updated 5 years ago by Xxxxxxx

the instructors' answer, *where instructors collectively construct a single answer*

I agree with Vyushita's answer. Also, mixing equations are linear whereas predator-prey equations are not.

Updated 5 years ago by Slobodan Simic

followup discussions *for lingering questions and comments*

Anonymous 5 years ago Are we going to see these on the exam since they weren't on the review?

Slobodan Simic 5 years ago Sorry, but I don't answer questions asking what types of questions will be on the final.

Reply to this followup discussion

Start a new followup discussion

Compose a new followup discussion

Xxxxxxx answered gggg in 3 min. 5 months ago

Figure 3: Snapshot of the Piazza forum

them into groups, in order to help instructors in focusing on specific areas in the course content. However, we draw both question-answer pairs and notes with a two-fold goal. First, we aim to provide the instructors with an overview of the forum activity, to draw their attention to topics that students struggled most with. Second, we also aim to provide the students a ready–reckoner for the course. Instead of waiting indefinitely for an instructors' response, they can leverage previous instructor curated or instructor endorsed answers, which in turn, can improve their productivity significantly.

3. BACKGROUND

Many universities use the Piazza forum facility [22] for asynchronous online discussions. In order to provide context for the experiments below, we provide a brief overview of this tool.

Piazza is a Q&A web service for online discussions, where users can ask questions, answer questions, and post notes. Every post can be organized into pre-defined folders (e.g. assignment1, logistics, etc) for easier organization. The user interface contains a dynamic list of posts, which are question titles followed by a snippet of lines from the post. The posts are arranged chronologically on the left panel, while the central panel is meant for viewing posts and adding contributions. For every question, there is a placeholder for the instructor's answer, which can only be edited by the instructors. There is also a students' answer section where students *collaborate* to construct a single answer. Students can *upvote* each other's questions or answers. Instructors can also *endorse* good questions and answers, which are then highlighted as instructor endorsed. There is also a discussion segment for follow-up threads. Figure 3 shows a snapshot of a class on the Piazza forum.

4. AUTOMATIC FAQ GENERATION

Our system addresses the problem of automatically identifying forum question–answer pairs and notes that are appropriate for inclusion in a course specific FAQ list.We worked on the creation of two separate FAQs, each for a different use case:

- *Students' FAQ* : This FAQ is intended to include mostly conceptual questions. Given that the key concepts covered in most classes change very little over the years, the concepts that students struggle with, have significant overlap over the years. Hence, this FAQ can increase their productivity by providing them instant help without having to wait indefinitely for an instructor's response. For the course staff the Students' FAQ translates to reduced work load.
- *Instructors' FAQ* : This FAQ is meant to serve as a snapshot of the previous offerings of the course. The FAQ is focused on the topics that the students struggled with, in the past, and on logistical mishaps of earlier offerings in order to help new or continuing instructors take corrective actions to improve the students' experience.

In the rest of the paper, we focus on the Students' FAQ. The Instructors' FAQ creation is similar to the Students' FAQ creation, with a slight shift in the focus of question-answer pairs and notes we intend to populate in each of them.

To obtain ground truth for training machine learning algorithms, human experts were consulted. A survey format was used, in which the experts were presented with a series of forum contributions (question–answer pairs and notes). The next section discusses how the authors chose the contributions to be used in the survey, and created the ground truth used for training.

4.1 Candidate FAQs and ground truth creation

We began with the complete dataset of contributions from four years of a graduate level course that introduces artificial intelligence (AI). This course had an average of 1610 questions from 2013 to 2016, with 2237 questions in the most recent dataset. The ratio of the total number of posts to the number of users on the forum was 10.95 in the most recent dataset. Given the large volume of the dataset, we filtered out questions that would be completely irrelevant from the FAQ perspective by using the following features.

- Number of upvotes on a question or note
- Number of upvotes on students' answer
- Number of upvotes on instructor's answer
- Number of unique collaborators in a thread
- Number of unique views
- Length of the follow-up thread

The reason for choosing the above simple features was to ensure that the system was generalizable for use with other similar forum datasets. Information related to upvotes, unique collaborators in the thread, unique views, number of follow-ups are readily available or derivable in most forum datasets. Besides, this was an easy way of indirectly crowd-sourcing the features.

Individual threshold percentiles were set for each of the above features. These thresholds served to assemble the candidate set for the expert opinion surveys. For instance, a threshold of 10^{th} percentile on the number of question upvotes would mean that, to be included in the survey, a question would need enough upvotes to clear that hurdle.

The thresholds for each of the above six features was set to 5$^{\text{th}}$ percentile for the most recent class offering. As we progressively selected candidate FAQs from older datasets, these thresholds were gradually increased. Our assumption was that more recent offerings would have posts that are more relevant to upcoming offerings. Our strategy was to be conservative and keep the individual filters very low in order to cover the entire breadth of potentially interesting questions. However, only the questions that crossed the thresholds for *all* the features survived. Apart from these features, we also relied on the folders/ tags in the Piazza dataset to filter out logistics related questions from the question pool for the Student FAQ. We accumulated the candidate FAQ set using four years of forum data for the same course. 62 question-answer pairs and notes were then randomly sampled from this cumulative candidate FAQ set.

Beyond the 62 posts, 13 additional random samples that did not cross the thresholds were included in the survey. This simple filter (using threshold percentile) narrowed our candidate set by 79.08% We found support that the excluded posts were justifiably elided, by observing the decisions of our experts. The experts selected none of the randomly sampled posts outside the candidate FAQ set, for inclusion in the FAQ. In total, 75 question–answer pairs and notes were selected for the survey and presented to the experts in random order.

In order to avoid fatiguing our experts, the set of 75 items was partitioned into three batches. These surveys were administered such that each question–answer pair or note was seen by at least three experts. All the experts were recruited from among the current course instructors for the same class from which the forum contributions were drawn. The survey instructions and one sample entry from the survey are included in the Appendix.

After examining the binary expert judgments, we found that their raw scores exhibited low agreement. In one of the groups, Expert 3 often voted in discordance with Expert 1 and 2; Experts 1 and 2 were more often in agreement. We focused in this initial work on strong evidence for an inclusion decision. We therefore relied only on judgments that were based on agreement between Expert 1 and Expert 2. This left us with a set of 54 judgments. Of these, 25 voted (unanimously) for inclusion of a post, and 29 voted against.

We compared two classification mechanisms for our task of deciding for, or against inclusion of a post in a FAQ: logistic regression, and random forest. We present both methods in turn. Our small set of human judgments forced us to rely on the intrinsic randomization of 10-fold cross validation repeated ten times. The random selection of predictor subsets, as trees are constructed, adds additional guard against overfitting. We found the predictions for the random forest model on the full sample set to be overly optimistic. Hence, we will leave the random forest model's full validation to a future step.

4.2 Logistic Regression Classifier

After centering and scaling, we allowed the R *glmnet* training to find an optimal LASSO lambda via grid search. The optimal lamda was computed as 0.1. The Prediction quality using the logistic regression classifier is shown in Table 1.

Table 1: Prediction quality statistics for the logistic regression classifier

Accuracy	**0.6852**
95% CI	(0.5445, 0.8048)
No Information Rate	0.537
P-Value[Acc > NIR]	0.019289
Kappa	0.3396
Mcnemar's Test P-Value	0.000685
Sensitivity	0.9655
Specificity	0.3600
Pos Pred Value	0.6364
Neg Pred Value	0.9000
Prevalence	0.5370
Detection Rate	0.5185
Detection Prevalence	0.8148
Balanced Accuracy	0.6628

Table 2: Confusion matrix of the logistic regression predictor

	Exclude	**Include**
Exclude	28	16
Include	1	19

Note that this classifier tends to be conservative, as evidenced in the number 16 in the confusion matrix. The classifier often tended to exclude forum contributions from the FAQ that the human experts included.

4.3 Random Forest Classifier

Using 10-fold cross validation, repeated 10 times We constructed a random forest classifier of 8000 random trees and allowed the statistics software to determine an optimal *mtry*, which is the number of randomly selected predictors to be used in each tree. Figure 4 shows the relationship between choices of this hyper parameter and accuracy. As reflected in the chart, the optimal $mtry == 2$. Figure 6 shows that the error rates stabilize as the number of trees increase. While we used an 8K tree model for our experiments, we observe that a 4K tree model would also suffice.

The OOB estimate of training error rate for an 8K-random forest is **31.48%**. Table 3 shows the confusion matrix for the trained model. As evidenced in the confusion matrix, we still need to improve the false positive rate of the model, before we proceed to test the classifier's performance.

Figure 5 shows the absolute decrease in accuracy if each of the predictors were removed from use in the classification.

Table 3: Confusion matrix of the trained random forest model

	Exclude	**Include**	**Class error**
Exclude	23	6	0.2068966
Include	11	14	0.4400000

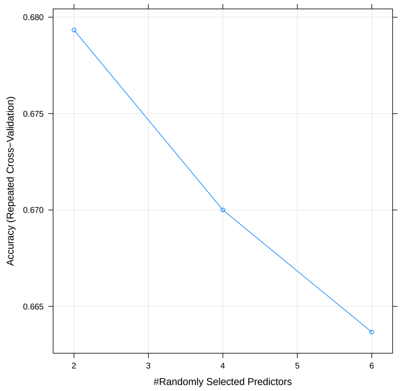


Figure 4: Change in the accuracy with number of randomly selected predictors

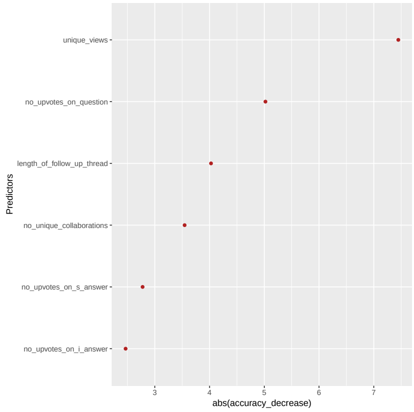


Figure 5: Absolute decrease in the accuracy on removing specific predictors

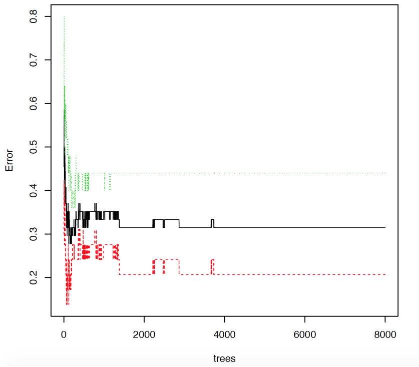


Figure 6: Number of trees for the random forest model. The black line indicates the out-of-bag error rate. The red line indicates the class error when predicting exclusion from the FAQ. The green line shows the inclusion error.

The chart is sorted such that the highest predictor on the vertical axis is the most important, as it contributes most effectively to the decisions. Note that the low position of `no_upvotes_on_i_answer` and `no_upvotes_on_s_answer` is not entirely reliable. This is because our candidate FAQ set included both questions and notes. In Piazza, notes do not have instructor's answer or students' answer segment. Thus, both these values, `no_upvotes_on_i_answer` and `no_upvotes_on_s_answer` were set to 0, which was the most frequent value for these two measures in our sample. The position of these predictors might have been higher if all posts had instructor–, and student answer sections.

Notice the high placement of *unique views*. In retrospect, views are the lowest friction method for students to 'vote with their eyeballs', which then manifests strongly in the classifier.

5. FUTURE WORK

Topic analysis of MOOC discussion content using LDA [6] has been explored recently in [7] and [5]. Given the promising results of LDA for similar tasks, we plan to use LDA for organizing our FAQ into topic–clusters, in order to make it easy for the end–users to search for questions of their interest. These topic clusters will also help in tag cloud generation. In [17], the authors present tag clouds as an interpretable representation mechanism and as a way to improve navigation and learning through the domain of knowledge. Using the LDA topic clusters, we plan to plot the top k words for each of the clusters in an interactive tag cloud, where each keyword or topic leads the user to a list of question-answer pairs in descending order of relevance.

In the near future, we plan to deploy FAQtor in a real world class and use the usage reports in order to improve FAQtor, similar to the approach in [14]. By allowing students and instructors to vote entries in and out of the FAQ, we plan to capture more patterns and continuously improve the FAQ.

As we gather more human judgments, we plan to test the Random Forest model's accuracy. We also plan to explore neural network approaches to include features that leverage the content of the posts as well.

Another line of follow-on research needs to be an investigation into the generalizability of our results to non-science classes. The success of such transfer learning is not guaranteed, but is not out of the question.

6. CONCLUSIONS

We present an evolving system that aims to extract additional benefits from class forum activity. In this paper, we presented the technology required for one of the system's components: FAQtor. We demonstrate how machine learning algorithms, in conjunction with the available forum statistics, can reveal which question-answer pairs and notes are relevant and important for the future offerings of the course.

To create the ground truth for training our classifiers, we consulted human experts. However, since this is an expensive process, our ground truth set is very small. Our experiments would certainly benefit from additional training data.

Unlike many related works in the FAQ world, which assume the pre-existence of a knowledge base, we have shown a principled approach to creating a solid starting point for the FAQ knowledge base by training classifiers that try to mimic human experts. While the prediction validity of the Random Forest classifier must still be examined, the logistic regression classifier's accuracy of 0.68 is significant. Nonetheless, the resulting FAQ knowledge base will need to be fine-tuned, and continuously refined using the techniques outlined in our future work section.

We propose that the records from online forum activity are important investments of human cognition. Since this investment is expensive, it is to every stakeholder's advantage to maximize the return on that effort. The work presented here is a step towards such maximization.

7. REFERENCES

[1] A. Agrawal, J. Venkatraman, S. Leonard, and A. Paepcke. Youedu: addressing confusion in mooc discussion forums by recommending instructional video clips. 2015.

[2] M. Alavi. Computer-mediated collaborative learning: An empirical evaluation. *MIS quarterly*, pages 159–174, 1994.

[3] M. A. Andresen. Asynchronous discussion forums: success factors, outcomes, assessments, and limitations. *Journal of Educational Technology & Society*, 12(1):249, 2009.

[4] Anonymous. Quantyler : Apportioning credit for student forum participation. *Submitted for publication.*

[5] T. Atapattu and K. Falkner. A framework for topic generation and labeling from mooc discussions. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*, pages 201–204. ACM, 2016.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[7] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth. Unsupervised modeling for understanding mooc discussion forums: a learning analytics approach. In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 146–150. ACM, 2015.

[8] K. Hammond, R. Burke, C. Martin, and S. Lytinen. Faq finder: a case-based approach to knowledge navigation. In *Artificial Intelligence for Applications, 1995. Proceedings., 11th Conference on*, pages 80–86. IEEE, 1995.

[9] W.-C. Hu, D.-F. Yu, and H. C. Jiau. A faq finding process in open source project forums. In *Software Engineering Advances (ICSEA), 2010 Fifth International Conference on*, pages 259–264. IEEE, 2010.

[10] KhanAcademy. *KhanAcademy*, 2017.

[11] H. Kim and J. Seo. High-performance faq retrieval using an automatic clustering method of query logs. *Information processing & management*, 42(3):650–661, 2006.

[12] H. Kim and J. Seo. Cluster-based faq retrieval using latent term weights. *IEEE Intelligent Systems*, 23(2):58–65, 2008.

[13] R. M. Marra, J. L. Moore, and A. K. Klimczak. Content analysis of online discussion forums: A comparative analysis of protocols. *Educational Technology Research and Development*, 52(2):23, 2004.

[14] A. Moreo, M. Romero, J. Castro, and J. M. Zurita. Faqtory: A framework to provide high-quality faq retrieval systems. *Expert Systems with Applications*, 39(14):11525–11534, 2012.

[15] J. B. Pena-Shaff and C. Nicholls. Analyzing student interactions and meaning construction in computer bulletin board discussions. *Computers & Education*, 42(3):243–265, 2004.

[16] L. F. Pendry and J. Salvatore. Individual and social benefits of online discussion forums. *Computers in Human Behavior*, 50:211–220, 2015.

[17] M. Romero, A. Moreo, and J. L. Castro. A cloud of faq: A highly-precise faq retrieval system for the web 2.0. *Knowledge-Based Systems*, 49:81–96, 2013.

[18] M. Scardamalia and C. Bereiter. Computer support for knowledge-building communities. *The journal of the learning sciences*, 3(3):265–283, 1994.

[19] R. Sindhgatta, S. Marvaniya, T. I. Dhamecha, and B. Sengupta. Inferring frequently asked questions from student question answering forums.

[20] E. Sneiders. Automated faq answering: Continued experience with shallow language understanding. In *Question Answering Systems. Papers from the 1999 AAAI Fall Symposium*, pages 97–107, 1999.

[21] S. D. Whitehead. Auto-faq: An experiment in cyberspace leveraging. *Computer Networks and ISDN Systems*, 28(1-2):137–146, 1995.

[22] Wikipedia. *Piazza*, 2017.

[23] C.-Y. Yang. A semantic faq system for online community learning. *JSW*, 4(2):153–158, 2009.

APPENDIX

A. SURVEY

The survey instructions for ground truth collection were:
Please choose 'Yes' or 'No' for each of the following questions/notes.
A 'Yes' indicates that the respective forum item would be relevant and useful in a FAQ list for the class.
A 'No' means that the item should not be included in the FAQ for future offerings of the class.
The entries in the FAQ are taken from previous iterations of the class. Their intent is to answer student questions instantly, and reduce TA workload (by answering common student questions)

One sample item from the survey is as follows:
Question: *For the definition of Markov Blanket does it refer to finding the neighbors of A in terms of the factor graph or in terms of the Bayesian network*
Response: *The Markov blanket refers to all variables that have a factor in common with at least one of the variables in A. If two variables are connected in a Bayesian Network, then they must share a factor in the corresponding factor graph representation.*