YouEDU: Addressing Confusion in MOOC Discussion Forums by Recommending Instructional Video Clips

Akshay Agrawal
Stanford University
akshayka@cs.stanford.edu

Jagadish Venkatraman
Stanford University
vjagadish@cs.stanford.edu

Andreas Paepcke
Stanford University
paepcke@cs.stanford.edu

ABSTRACT

In Massive Open Online Courses (MOOCs), struggling learners often seek help by posting questions in discussion forums. Unfortunately, given the large volume of discussion in MOOCs, instructors may overlook these learners' posts, detrimentally impacting the learning process and exacerbating attrition. In this paper, we present YouEDU, an instructional aid that automatically detects and addresses confusion in forum posts. Leveraging our publicly-available Stanford MOOCPosts corpus, we train a heterogeneous set of classifiers to classify forum posts across multiple dimensions. In particular, classifiers that target sentiment, urgency, and other descriptive variables inform a single classifier that detects confusion. We then employ information retrieval techniques to map confused posts to minute-resolution clips from course videos; the ranking over these clips accounts for both video-clickstream data and textual similarity between posts and closed captions. We measure the performance of our classification model in multiple educational contexts, exploring the nature of confusion within each; we also evaluate the relevancy of materials returned by our ranking algorithm.

Keywords

ACM proceedings, LaTeX, text tagging

1. INTRODUCTION

* Proliferation of MOOCs
* Volume of posts high
* Difficult to get a birds-eye view of the course, difficult to address it.

* Work looking into sentiment thus far is limited by datasets
* Work has been done on confusion, but not so much on MOOCs (save RosÃl')
* Work into intelligently intervening + aiding the instructor
* Previous work has found forum to perhaps not be the most useful, even

* We suspect that the forum's perceived lack of usefulness is not instrinsic but rather lack of attention lack of instructor tools + isolation from other parts of the classroom.
* We accordingly set out to adress both of these problems âĂŤ mining for affect gives instructors a pulse on the state of the course, and linking to videos marries forum and other course resources.

The remainder of this paper examines related work (section two), presents the Stanford MOOCPosts corpus (section three), sketches the architecture of YouEDU (section four), details its constituent classification and recommendation phases, evaluating both and interpreting results (sections five and six), and proposes future work (section seven).

2. RELATED WORK

3. THE STANFORD MOOCPOSTS CORPUS

A precondition to automatically detecting affect in MOOC discussion forums was manually identifying it; given that no publicly-available corpus of tagged MOOC discussion forum posts existed prior to our research, we set out to create our own. The outcome of our data compilation and curation was the Stanford MOOCPosts dataset: a corpus composed of 29,604 anonymized learner forum posts from eleven Stanford University public online classes. Freely available to academic researchers, the MOOCPosts dataset was designed to enable computational inquiries into the nature of both affect and content in MOOC discussion forums.

Each post in the MOOCPosts dataset was scored across six dimensions – confusion, sentiment, urgency, question, answer, and opinion – and subsequently augmented with additional metadata. In this section, we detail the data collection methodology, defining each of the six dimensions along the way, and briefly present some insights gleaned by mining the set.

3.1 Methodology: Compiling the Dataset

Nine judges from oDesk were hired to ...

3.2 Insights and Discussion

We report insights gleaned into the nature of affect, etc. across these courses.

3.2.1 Relationship between Variables

In this section, we report the pairwise correlations between variables to 1) shed some light into the nature of each and also 2) to motivate a YouEDU design choice.

4. YOUEDU: INTELLIGENT INTERVENTIONS

As the name might imply, YouEDU is an intervention system with granularity at the level of the individual ...

5. PHASE I: DETECTING CONFUSION

In this section, we present the model used to classify confusion. At a high-level, we use a coordinated combination of classifiers ...

5.1 Classifier Design

5.1.1 Feature Space

Here's an intuition of what we figured might be helpful across all variables ...

* Bag of Words
(* Specific Bigrams / words / trigrams? Time permitting)
* Pre-processing Steps
* Features Extracted / Generated

5.1.2 A Combination of Hypotheses

* Classifier Combination – Overview
* Our particular implementation
* Subclassifiers Variable-Specific Features
* Training, use gold values, testing, use predicted values
* Combination Step Logistic Regression Layer

5.2 Evaluation

* educational contexts
* metrics used
* results
* implications

6. PHASE II: MAPPING CONFUSED POSTS TO VIDEO CLIPS

6.1 The Recommendation Algorithm

6.1.1 Retrieval

6.1.2 Ranking

6.2 Evaluation

Two experts were hired ...

7. FUTURE WORK

Future work might focus on strengthening the link between the classifiers and the reccomendation system; in particular, it would behoove us to devise a way to filter our set of confused posts to a subset for which recommendation makes sense. Additionally, we might want to make our classifiers better and index back into the previous course to retrieve answers for courses. Deploying this system live is another thing that we might do.

8. CONCLUSION

YouEDU takes an initial step towards building automated confusion intervention ...

9. ACKNOWLEDGMENTS

This section is optional[1]; it is a location for you to acknowledge grants, funding, editing assistance and what have you. In the present case, for example, the authors would like to thank Gerald Murray of ACM for his help in codifying this *Author's Guide* and the **.cls** and **.tex** files that it describes.

10. REFERENCES

[1] M. Wen, D. Yang, and C. P. Rosé. Sentiment analysis in mooc discussion forums: What does it tell us? *Proceedings of Educational Data Mining*, 2014.