#lets-discuss: Analyzing Student Affect in Course Forums Using Emoji

Ariel Blobstein
Ben-Gurion University
arielblo@post.bgu.ac.il

Kobi Gal
Ben-Gurion University
University of Edinburgh
kobig@bgu.ac.il

Hyunsoo Gloria Kim
University of California, Davis
hyun@ucdavis.edu

Marc Facciotti
University of California, Davis
mtfacciott@ucdavis.edu

David Karger
MIT
karger@mit.edu

Kamali Sripathi
University of California, Davis
ksripathi@ucdavis.edu

ABSTRACT

Emoji are commonly used in social media to convey attitudes and emotions. While popular, their use in educational contexts has been sparsely studied. This paper reports on the students' use of emoji in an online course forum in which students annotate and discuss course material in the margins of the online textbook. For this study, instructors created 11 custom emoji-hashtag pairs that enabled students to quickly communicate affects and reactions in the forum that they experienced while interacting with the course material. Example reporting includes, inviting discussion about a topic, declaring a topic as interesting, or requesting assistance about a topic. We analyze emoji usage by over 1,800 students enrolled in multiple offerings of the same course across multiple academic terms. The data show that some emoji frequently appear together in posts associated with the same paragraphs, suggesting that students use the emoji in this way to communicating complex affective states. We explore the use of computational models for predicting emoji at the post level, even when posts are lacking emoji. This capability can allow instructors to infer information about students' affective states during their "at home" interactions with course readings. Finally, we show that partitioning the emoji into distinct groups, rather than trying to predict individual emoji, can be both of pedagogical value to instructors and improve the predictive performance of our approach using the BERT language model. Our procedure can be generalized to other courses and for the benefit of other instructors.

Keywords

Affect recognition, Course Forums

1. INTRODUCTION

Students typically experience a range of affective responses to learning materials, taking the form of emotions, beliefs

and attitudes that have a profound effect on their learning (and vice versa) [7]. Understanding what students are feeling when they learn can, therefore, be useful for teachers who may wish to adapt course content and/or delivery in response to student affect [26, 18, 1].

Emoji are conventionally used in social media to enhance the meaning of text or to be used as substitute for words [17]. When conversing through text, such as with written posts in course forums, emoji can be a natural way for students to express what they are feeling about the learning material. A body of work in the social and computational sciences has studied people's use of emoji, and used them to facilitate computational tasks such as sentiment analysis and emotion recognition [19, 13]. However, the use of emoji by students in educational forums has been much less studied. Our contribution is a computational analysis of students' use of emoji in a course forum with over 1,800 of students.

Students in this course use the Nota Bene (NB) collaborative annotation-based forum, which allows them to anchor messages directly to reading material in their online course textbook. Course instructors designed a set of 11 emoji that allow students to communicate affective responses in their posts, such as expressing curiosity or confusion about topics in the course, or inviting discussion about a topic. Students were incentivized to use these emoji by receiving course points on their reading assignment for using at least one emoji in their forum posts.

We provide a detailed analysis of emoji usage in the course at the post level as well as the paragraph level in the reading material that the post refers to. We find that the most commonly used emoji request discussion and assistance from instructors or peers, and that in some cases students combine several emoji in the same post to create complex affects. Some emoji frequently appear together in posts that are anchored in the same paragraph, suggesting that students may simultaneously experience a combination of responses to the course readings or that they choose to use multiple emoji to express more complex affects that are not sufficiently well represented by a single emoji.

We explore the use of computational models for classifying emoji use at the individual and group level. We use a

pre-trained deep-learning language model (BERT) to predict emoji at the post level. This model significantly outperforms a Long-Short-Term-Memory (LSTM) architecture that was used by others to predict emoji in social media model, but was trained on the same task. [13, 3].

In agreement with course instructors, we cluster the emoji into categories, based on the relationship between the predicted emoji probabilities that are outputted by the BERT model. We partition the emoji into eight distinct groups (3 groups containing an emoji pair, and 5 groups containing a single emoji). Partitioning the emoji into these distinct groups makes pedagogical sense to the course instructors while also improving performance when predicting emoji-groups rather than individual emoji.

The contribution of this work is in a new computational model for predicting student affect that was trained on students' self-reported emoji without requiring hand-labeling of the data by experts. The model can directly be applied to any discussion forum, even where emoji information is not readily available, potentially increasing its impact. Our work can help instructors, particularly in high-enrollment courses, decide where and how to intervene in discussions, and to assess where content or course revisions might be made to improve student learning.

2. RELATED WORK
Our work relates to past studies for inferring student affect from their online text conversations, as well as computational work that analyzes the use of emoji in social media. We mention relevant prior works below and refer the reader to the review of the field by Kastrati et al. [16] for additional details.

Basic approaches to infer students' sentiments in online conversations used classic NLP methods (parsing, lexical dictionaries) [8, 6]. Studies showing the correlation between affects and dropout [28] and the correlation between affects and exams scores exist [23]. Jena [15] used classical machine learning methods (SVMs, Naive Bayes) to learn the sentiment polarity (positive, negative, and neutral) from students' posts, as well as predicting basic students' emotions from text (e.g., anxiety, bored, confused, excited). Estrada et al. [12] used deep neural networks, such as convolutional neural network and LSTMs, as well as evolutionary generative models, in order to classify sentiment polarity and emotions.

As noted by Kastrati et al. there are relatively few works on recognizing students' emotions from text, despite the pedagogical importance of this task, and the growing prevalence of online learning [26, 18]. One possible reason for this scarcity of works is the reliance on hand-labeled data sets which are costly and time-consuming to obtain. We directly address this gap in our work by using students' self-reported affect in the form of emoji as proxies for their emotional state.

There is growing research studying the use of emoji as tools in computer-mediated communication [25, 17]. Emoji can be used in two different ways. They can appear alongside text to include an affect or to enhance the meaning of a text. They can also be used as a substitute for words, or to conform a new notion.

Zhang et al. [29] investigated the use of emoji among students in the Nota Bene framework and showed the potential of detecting students' affects by training a classifier in order to distinguish between confusion and curiosity. Geller et al. [14] defined rules for confusion detection that are based on students' use of two types of emoji. They showed that the resulting rules closely align with the ground truth judgement of educational experts. We generalize both works to the more challenging task of recognizing multiple affects, and explore ways to facilitate this task by combining clustering methods with input from the course staff.

Felbo et al. [13] trained an LSTM architecture to predict emoji use in Twitter as proxies for users' emotional states. They show the model was able to generalize to other datasets containing self-reported emotional states. They employed clustering methods to learn relationships between 64 different emoji. We go beyond this work in several ways. First, we study the relationship between emoji use on the topic level. Second, we use the clusters to build better predictive models. Third, we involve course instructors in the analysis and use them to determine the best partition.

Çöltekin and Rama [10] used SVMs to predict emojis with a bag-of-b-grams feature set, combining both character n-grams and word n-grams and weighted by the TF-IDF score. This model achieved top performance in a recent competition for predicting 20 emoji on Twitter (SemEval 2018 task 2) [2]. Zhang et al. [30] used a BERT model for the emoji prediction task that outperformed Çöltekin and Rama [10] approach. We directly extend Zhang et al. model in adapting BERT to a biology course setting with an additional pre-training over the course's previous data.

3. THE NB SETTING
The Nota Bene (NB) web application is an open source social annotation tool that was developed at MIT [31]. NB is used in hundreds of university courses and includes more than 40,000 registered student users. The main feature of Nota Bene gives users the ability to directly annotate course content. Course content (PDF, HTML, or video file) is uploaded to the NB website by instructors. Students can annotate the content by highlighting a passage in the document (called "the marked text") and then add a post by typing into a text field that appears in the margin. These annotations may be used to create a post or to ask questions about the content. Classmates are encouraged to reply to other students' comments and to answer any posted questions.

NB posts are organized into threads, which consist of a starting comment or question followed by all the replies made by other students or instructional team to the initial annotation or to the subsequent replies. The in-place structure of the NB tool allows students to interact in the forum while they are reading the course material and provides context to the discussion. This structure has been shown to be beneficial for learning [22]. The NB interface allows students to express emotions and other affects in their comments via hashtags (which are translated to graphical representations of emoji), thus allowing students and instructors to filter

students' comments based on the type of affect in which they are interested.

3.1 The FYBIO course and emoji

FYBIO (First Year BIOlogy) is a general biology course required for all life sciences majors at a large, public university and is typically taken by students during their first year of study. Depending on the academic term it is offered, the course consists of 25 or 26 lectures. During this study, the course staff posted reading materials before each lecture and the students were assigned to read these materials and to provide three substantial posts in NB before each class. This encouraged active participation in forum discussions. Students received additional credit for including at least one emoji in at least one of their posts per lecture. The NB interface displays the hashtags graphically using relevant emoji symbols in students' posts. The emoji were designed by the course instructors to allow students the option to express emotions and opinions about course material, as well as invite assistance or participation from their peers or instructors. For the remainder of this paper, we will use the term "affects" to refer to all of the above expressions.

A full list of the emoji and their intended uses is shown in Table 1. Some examples of emoji and their intended uses are as follows: The #i-think emoji expresses an idea to share; the #lets-discuss emoji invites students and instructors to contribute to a discussion; the #learning-goal emoji identifies a topic or idea related to the course objectives; the #question emoji requests help from the course staff regarding a topic.

Emoji	Associated hashtag	Intended use	Percentage of use
?	#question	…you would like the professor to address a specific question in class.	18.7%
💭	#i-think	…you have an idea to share about something but are not sure if it is accurate.	18%
🤔	#just-curious	…you would like to learn more about a certain topic that you are curious about or intrigued by.	13.4%
‼️	#important	…if you found a topic or idea to be important.	11.7%
♥	#interesting-topic	…you found a topic or idea to be interesting.	10.7%
💡	#lightbulb-moment	…something (e.g., course material, a peer's post) led you to a new understanding (an "ah-ha" moment).	8.3%
💬	#lets-discuss	…you would like to invite your fellow classmates and instructors to weigh in, in a NB discussion.	5.6%
🌍	#real-life-application	…a topic or idea has any implications in the real world.	4.2%
😲	#surprised	…you found a topic or idea to be surprising.	3.9%
😵	#lost	…you are totally confused or overwhelmed about a topic and have no idea what question to even ask.	2.9%
🖼	#learning-goal	…you identify a topic or idea related to the course objectives or reading-specific learning goals	2.5%

Table 1: Hashtags in NB, their associated emoji, intended uses and percentage of usage

Nota Bene provides instructors with a heatmap showcasing students' use of the different emoji (See Figure 1). Each emoji is displayed using a different color, and the brightness varies with respect to the density of the emoji in the posts.
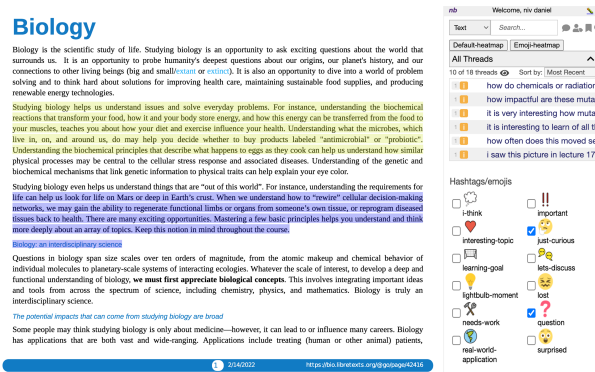
Instructors can filter which emoji to display.

Figure 1: Emoji heat map in Nota Bene

Table 2: Statistics of our data set (different course instances of FYBIO)

Statistics (Num.)	Winter 2021	Summer 2021	Winter 2022	Total
Lectures	14	26	14	**54**
students	758	182	905	**1,845**
posts	28,072	15,072	36,165	**79,309**
emoji	23,867	10,635	28,674	**62,906**

Our dataset contained 79,309 unique student posts from three instances of the FYBIO course, as shown in Table 2. In total, 55,437 posts contained at least one emoji.

All students who participated in FYBIO courses filled a consent form allowing their data to be anonymized and analyzed for the sake of this study. The study was reviewed by the IRB (1456274-1) and deemed exempt. All students whose data was included in the study filled out a consent form and opted into the study.. The high fraction of posts containing emoji (over 70%) shows that students' use of this tool went beyond the requirement in the course and echos past work demonstrating the educational benefits of affect [1, 18, 27].

4. METHODOLOGY AND RESULTS

Our goal is to provide teachers with better tools for making sense of students' affect in the course. To this end, we explore two main research questions. First, how do students use emoji in their posts? Second, can we use machine learning to infer how students would have tagged posts using the available emoji?

Our methodology addresses the first research question with a detailed analysis of emoji usage at the post and paragraph level, and addresses the second research question with the design of language models for emoji classification from text.

4.1 Analysis at the post level

Table 1 shows the percentage of emoji use in FYBIO. The table shows that the #question and #i-think emoji were used most often. Both of these emoji reflect uncertainty about the material and invite participation by instructors and students. The emoji expressing more direct requests for participation (e.g., #lets-discuss, #lost) were used much less frequently. This may reflect a resistance towards revealing

information about understanding that may have an adverse affect for students (e.g., peer pressure or getting a lower grade). Future editions of FYBIO will allow anonymous posting, which may change the way students use these posts.

By analyzing students' comments we have found that about 70% of posts included at least one emoji. The majority (about 89%) of these posts include a single emoji, showing that students express a single affect in each post. The use of multiple emojis in the same post creates a complex affect. 10% of the posts contained two emoji and less than 1% of the posts contained 3 or 4 emoji.

4.2 Classifying individual emoji

In this section we describe how we use language models to classify students' use of emoji at the post level. A good classification model can potentially aid instructors in making sense of students' affective states in situations where emoji are not used, which is the case for most forums.

The model provides a mapping from a post to the most relevant emoji for the post according to the family. To this end we define a multiclass prediction task where the target class is the set of emoji shown in Table 1.

An instance was created for each post in the FYBIO dataset, and labeled with the relevant emoji for the post. For posts with multiple emoji, duplicate instances of the post were created for each emoji. The instances were split into training (85%) and test sets (15%), such that 15% randomly sampled instances of the training set are used as a validation set. We stratified the training and test sets such that each set contains approximately the same percentage of samples of each target class. In no case did the training and test sets include an instance with the same post. We compare two types of language models for the classification task. BERT is a pre-trained, transformer based language model that is commonly used in state-of-the-art natural language tasks [11]. We used an open-source BERT configuration ("bert-base-uncased") that was trained on broad domain corpora (English Wikipedia and BooksCorpus).[1] The architecture contains 12 layers, 768 hidden units and 12 attention heads. We use this architecture in all of our experiments. Augmenting BERT with a pre-training procedure has demonstrated promising results in downstream NLP tasks [5, 20].

Our proposed model architecture (See Figure 2) consists of two components: 1) A language representation model (pre-training step) and 2) a fully connected neural network (fine tuning step). The language representation model was pre-trained using two sources: The FYBIO text book (150K words), as well as students' posts from past instances of the course in 2020 (6.4M words). "bert-base-uncased" comes with a predefined vocabulary which consist of the approximately $30K$ most frequent words and sub-words from its pre-training corpus. We initiated the BERT model previously described, with its pre-trained weights, and afterwards trained the model for 50 additional epochs with a learning rate of $2e - 5$, which gave us the best results. Note that BERT allows the user to add special words or sub-words

---

[1]`https://huggingface.co/transformers/modeldoc/bert.html`

that are unique to our domain. Since we pretrained using the Biology text book, we did not encounter words that are out of vocabulary, and so did not change the vocabulary.

The neural network was trained for the emoji classification task using 768 nodes with a softmax activation function above the additional pre-trained model. The input to this network was the embedded student's post using the language representation model. The output of the network is a probability distribution over the 11 emoji in the target class. We fine-tuned the model with an addition of three epochs with a learning rate of $2e - 5$, using a maximal sentence length of 250, which gave us the best results.

We compared BERT to a bi-directional LSTM architecture, similar to the one used by Felbo et al. and Baziotis et al. [13, 4] to classify emoji in social media. The model included five layers. One layer was used to embed words in students' posts as high-dimensional vectors; three layers for classification consisting of 64 LSTM units (32 units in each direction); the final layer was an attention mechanism layer that connects words in posts with preceding and succeeding words while computing the importance of each word with the corresponding label. The model was implemented using Python's keras package [9]. We used a separate pre-training process using word2vec [21] to construct a high dimensional 200-sized vector representation of words in students' posts. The pre-training used the FYBIO textbook, as well as students' posts from 2020 and implemented using Python's Gensim package [24].

Table 3 compares the BERT and LSTM models when classifying emoji according to precision, recall, and weighted F-1 scores. We can see that the BERT model outperforms the Bi-LSTM model in all three metrics by a significant margin (McNemar's test, $p < 1.16 \cdot e^{-35}$). We attribute this difference to the pre-trained language model in BERT which allows it to generalize to domains with low amounts of training data [11].

Table 3: Model output comparison

Model	Precision	Recall	F-1 Score
BERT	**40.2%**	**43.3%**	**40.7%**
LSTM	26.6%	32.8%	29.3%

We note that prior work using BERT to classify emoji report a macro F-1 score of 38.5% [30], while our Macro-F1 score is lower (32.2%). However, we do not compare directly with these models for several reasons. First, they used an order of magnitude more data (550K tweets vs. 50K posts). Second, their target set was larger (20 vs. 11 emoji). Third, the emoji setting is different (e.g., smiley-faces, hearts, etc.) and is used in different ways (e.g., use the same emoji in different parts of the sentence).

Table 4 breaks down the performance of the model according to individual emoji. The table shows a positive relationship between the amount of training data for a given emoji and the prediction performance of the model for the emoji. An interesting exception is the #lets-discuss emoji, which is pedagogically important and represents more than 6% of the dataset, but achieves very low performance. The reason
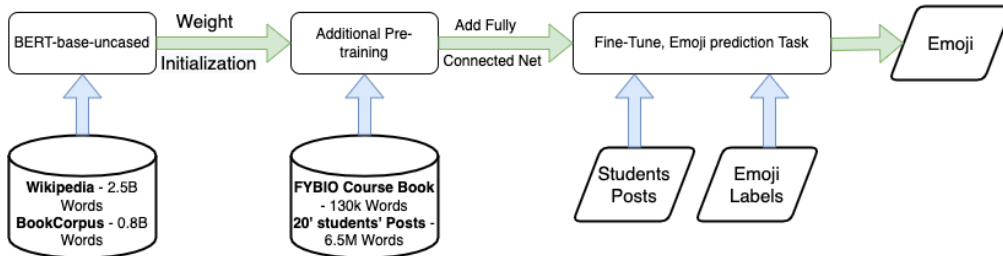
Figure 2: Model architecture used for Emoji prediction

Table 4: BERT prediction score, divided by prediction output score for each emoji

Metric	❓	💭	🤔	‼️	❤️	💡	💬	🌍	😯	😵	🥅
Precision	0.55	0.49	0.37	0.4	0.33	0.33	0.18	0.31	0.33	0.39	0.19
Recall	0.72	0.62	0.28	0.52	0.34	0.26	0.05	0.23	0.28	0.2	0.05
F-1 Score	0.62	0.55	0.32	0.45	0.34	0.29	0.08	0.26	0.3	0.26	0.07

is that the model misclassified over 50% of #lets-discuss as #i-think, suggesting that students may use these two emoji interchangeably to convey the same affect. We study this question in the next section.

Another possible reason for the low performance for some of the emoji may be that they were used by students to convey different affects than those intended for the emoji. For example, the following post shows a student has reached a conclusion but does not understand why it is true. This reflects a question rather than total confusion which was the originally intended purpose of this emoji: *I also get the impression that not having similar proof-reading mechanisms in transcription means that they are less severe. However. I'm not sure why this would be the case #lost.*

4.3 Classifying groups of emoji

In this section, we exploit the fact that emoji are often used together in paragraphs to cluster the emoji into pedagogically meaningful categories and predict the categories, rather than individual emoji. Felbo et al [13] clustered emoji based on the relationships between predicted probabilities of a computational model. We follow their approach and apply a hierarchical clustering algorithm with average linkage on the correlation matrix on the predicted probabilities of the BERT model shown in Table 4. Each emoji is represented as a vector of predicted probabilities for each post in the test set. To illustrate this approach, consider the following post from the test set: *is there an operon for every single metabolite? [trp] operon applies to tryptophan while lac operon applies to lactose. how many different types of operons are there and how do they function differently?* The BERT model, applied to this post, outputted the highest probability for the #question and #just-curious emoji (0.467 and 0.45 probability, respectively) and low probability for the rest of the emoji (e.g., 0.008 probability for the #interesting-topic emoji, and 0.003 probability for the #lightbulb-moment emoji). Note that because we are using the softmax function, these probabilities add up to 1, although the two emoji have a combined probability of approximately 0.92. The fact that the model is likely to assign a high or low probability to both #question and #just-

curious emoji for a given post suggests that they should form a single category.

Figure 3 shows the dendrogram that is outputted by applying hierarchical clustering. The height of each node in the y-axis is proportional to the value of the intergroup dissimilarity between its child nodes. The distance threshold of the clustering is used to determine which nodes to put in the same cluster.

In concurrence with the FYBIO instructions, we determined a threshold of 1. This threshold grouped the 11 emoji into three emoji pair clusters and 5 singleton emoji clusters: (#question, #just-curious),(#lets-discuss, #i-think), (#important, #learning-goal), (#surprised), (#lightbulb-moment), (#lost), (#interesting-topic), and (#realworld-application).
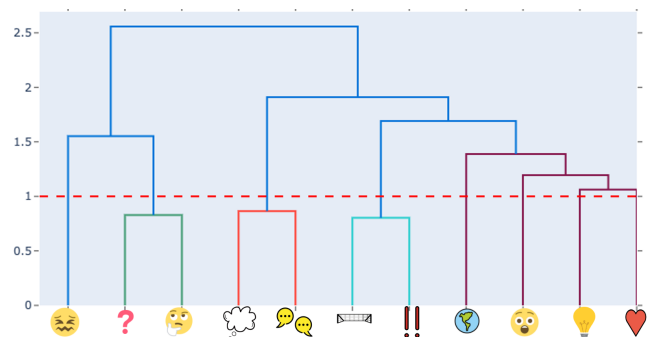

Figure 3: Hierarchical clustering of the emoji based on the model's prediction. The dashed red line represents the confidence threshold.

The justification provided by the course instructors for this division are as follows: With respect to the cluster containing the emoji (#question,#just-curious), instructors claimed that they invite a response from peers or the instructors, remarking that *"They are both requesting responses but with different urgency."* The individual emoji were also highly correlated on the paragraph level. Both of these aspects led instructors to agree to put them in the same category.

Table 5: Model score of each cluster and the weighted average score of all labels

Cluster	Precision	Recall	F-1 Score
#question, #just-curious	77.8%	83.9%	80.7%
#lets-discuss, #i-think	58.4%	60.5%	59.5%
#important, #learning-goal	47.3%	51.4%	49.3%
Total Score (8 Categories)	55.2%	56.8%	55.8%

With respect to the cluster containing the emoji (#lets-discuss, #i-think), instructors claimed that *"both of these emoji indicate enthusiasm to continue around a topic, either for curiosity or sometimes to clarify [...] they are expressing how they understand it and want clarification or alternate views from peers."* The individual emojis were also highly correlated on the paragraph level.

With respect to the cluster with the emoji (#important, #learning-goal), instructors claimed that both emoji depict students' perspectives on exam-related content. They claimed that *"these emojis are used when students identify parts of the text that they believe are important for them to perform well in the course - that may be linked to assessment or the development of knowledge that build towards good performance on assessments."*

Interestingly, the emoji-pair (#surprised, #interesting-topic) exhibited a high correlation on the paragraph level but was deemed sufficiently distinct by instructors to warrant a separate affect category for each emoji. Instructors claimed that "Both emoji are expressions of "enjoyment" of the text - though as noted slightly different."

The last part of our methodology was to study whether predicting the 8 categories, rather than the 11 original emojis, would improve prediction performance. To this end we trained the BERT model with the same architecture as described in Section 4.2, using the 8 categories as the target set. Table 5 shows the performance for the three emoji-pair clusters as well as total performance over all 8 of the categories. As shown by the Table, there was an improvement of 15% in prediction performance (from 40% to 55.8%) when compared to using a target set containing the original 11 emoji. We note that an improvement in prediction performance was to be expected, given the reduction in the size of the target set. What is interesting is that the two emoji #lets-discuss and #learning-goal, which achieved very low performance when predicted as individual emoji (see Table 4), exhibited significant improvement when clustered together with another emoji (F-Score of 59.5% and 49.3% respectively). The cluster with the emoji pair (#question, #just-curious) received the greatest score across all metrics. In terms of predicting the clusters with a single emoji, we saw a slight reduction in performance (less than one percent).

5. DISCUSSION AND CONCLUSIONS

This paper studied students' use of emoji in a large scale course forums used by hundreds of students. We found that when made available to them, students tend to use emoji in their posts to enhance their meaning or to express affect (an emotion, belief and opinion), rather than as a substitute for words. The most popular use of emoji was to invite further explanation or to express an interest in a given topic. Emoji expressing confusion or misunderstandings were less popular, and may reflect students' hesitation to expose these affects in public. Students' use of emoji is complex. In some cases, students use multiple emojis in the same post to convey a joint affect, while in other cases students use different emoji interchangeably to mean the same affect. Some students may use emoji in a different way than their intended use, and instructors may wish to present their meaning to students in the beginning of the course.

We began to explore the use of emoji as a pedagogical tool by instructors to aid course design or guide students. To this end, we analyzed how topics in the course material generate different emoji reactions from students. This led us to partition emoji into groups using hierarchical clustering. We identified the most pedagogically meaningful emoji clusters with the help of the course instructors, and designed a language model based on BERT that was able to classify students' posts to the right cluster with good performance.

The language model allows to classify students' posts with affects even when emoji are absent from the post, which can naturally extend our contributions to other courses and forums. By adding more data to the learning process, we hope to improve our language model. Following the improvement, we want to apply the model in two separate scenarios. (1) non-bio NB course (with same emojis) (2) course forum with different labels in hope that the model is able to learn affects signals from students writes.

We also intend to use our insights and computational tools to help teachers make sense of student affect in an active class. We envision a dashboard that would alert instructors to affects they care about, like confusion, or insights conveyed by students about learning goals. We wish to study how teachers use this tool to inform their course design, or to actively intervene in a forum to guide discussion.

Acknowledgements

6. ADDITIONAL AUTHORS

Jumana Almahmoud (MIT, email: `jumanam@mit.edu` and Michele Igo (U.C. Davis, email: `mmigo@ucdavis.edu`)

7. REFERENCES

[1] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4):223–241, 2010.

[2] F. Barbieri, J. Camacho-Collados, F. Ronzano, L. Espinosa-Anke, M. Ballesteros, V. Basile, V. Patti, and H. Saggion. SemEval 2018 task 2: Multilingual emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*,

pages 24–33, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[3] C. Baziotis, N. Athanasiou, G. Paraskevopoulos, N. Ellinas, A. Kolovou, and A. Potamianos. NTUA-SLP at semeval-2018 task 2: Predicting emojis using rnns with context-aware attention. *CoRR*, abs/1804.06657, 2018.

[4] C. Baziotis, N. Athanasiou, G. Paraskevopoulos, N. Ellinas, A. Kolovou, and A. Potamianos. Ntua-slp at semeval-2018 task 2: Predicting emojis using rnns with context-aware attention. *arXiv preprint arXiv:1804.06657*, 2018.

[5] I. Beltagy, K. Lo, and A. Cohan. Scibert: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676*, 2019.

[6] H. H. Binali, C. Wu, and V. Potdar. A new significant area: Emotion detection in e-learning using opinion mining techniques. In *2009 3rd IEEE International Conference on Digital Ecosystems and Technologies*, pages 259–264, 2009.

[7] M. Boekaerts and R. Pekrun. Emotions and emotion regulation in academic settings. In *Handbook of educational psychology*, pages 90–104. Routledge, 2015.

[8] D. S. Chaplot, E. Rhim, and J. Kim. Predicting student attrition in moocs using sentiment analysis and neural networks. In *CEUR Workshop Proceedings*, volume 1432, pages 7–12, 2015.

[9] F. Chollet et al. Keras. https://keras.io, 2015.

[10] Ç. Çöltekin and T. Rama. Tübingen-oslo at semeval-2018 task 2: Svms perform better than rnns in emoji prediction. In *Proceedings of The 12th International Workshop on Semantic Evaluation*, pages 34–38, 2018.

[11] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[12] M. L. B. Estrada, R. Z. Cabada, R. O. Bustillos, and M. Graff. Opinion mining and emotion recognition applied to learning environments. *Expert Systems with Applications*, 150:113265, 2020.

[13] B. Felbo, A. Mislove, A. Søgaard, I. Rahwan, and S. Lehmann. Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm. *arXiv preprint arXiv:1708.00524*, 2017.

[14] S. A. Geller, N. Hoernle, K. Gal, A. Segal, A. X. Zhang, D. Karger, M. T. Facciotti, and M. Igo. # confused and beyond: detecting confusion in course forums using students' hashtags. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 589–594, 2020.

[15] R. Jena. Sentiment mining in a collaborative learning environment: capitalising on big data. *Behaviour & Information Technology*, 38(9):986–1001, 2019.

[16] Z. Kastrati, F. Dalipi, A. S. Imran, K. Pireva Nuci, and M. A. Wani. Sentiment analysis of students' feedback with nlp and deep learning: A systematic mapping study. *Applied Sciences*, 11(9):3986, 2021.

[17] L. Kerslake and R. Wegerif. The semiotics of emoji: the rise of visual language in the age of the internet (book review). *Media and Communication*, 5(4):75–78, 2017.

[18] B. Kort, R. Reilly, and R. W. Picard. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE international conference on advanced learning technologies*, pages 43–46. IEEE, 2001.

[19] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič. Sentiment of emojis. *PLoS ONE*, 10(12):e0144296, 2015.

[20] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4):1234–1240, 2020.

[21] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[22] K. Miller, S. Zyto, D. Karger, J. Yoo, and E. Mazur. Analysis of student engagement in an online annotation system in the context of a flipped introductory physics class. *Physical Review Physics Education Research*, 12(2):020143, 2016.

[23] Z. A. Pardos, R. S. Baker, M. O. San Pedro, S. M. Gowda, and S. M. Gowda. Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, 1(1):107–128, 2014.

[24] R. Rehurek and P. Sojka. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.

[25] G. Santamaría-Bonfil and O. G. T. López. Emoji as a proxy of emotional communication. In *Becoming Human with Humanoid-From Physical Interaction to Social Intelligence*. IntechOpen, 2019.

[26] R. Sylwester. How emotions affect learning. *Educational leadership*, 52(2):60–65, 1994.

[27] G. J. Trevors, K. R. Muis, R. Pekrun, G. M. Sinatra, and M. M. Muijselaar. Exploring the relations between epistemic beliefs, emotions, and learning from texts. *Contemporary Educational Psychology*, 48:116–132, 2017.

[28] M. Wen, D. Yang, and C. Rose. Sentiment analysis in mooc discussion forums: What does it tell us? In *Educational data mining 2014*. Citeseer, 2014.

[29] A. X. Zhang, M. Igo, M. Facciotti, and D. Karger. Using student annotated hashtags and emojis to collect nuanced affective states. In *Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale*, pages 319–322, 2017.

[30] L. Zhang, Y. Zhou, T. Erekhinskaya, and D. Moldovan. Emoji prediction: A transfer learning approach. In K. Arai, S. Kapoor, and R. Bhatia, editors, *Advances in Information and Communication*, pages 864–872, Cham, 2020. Springer International Publishing.

[31] S. Zyto, D. Karger, M. Ackerman, and S. Mahajan. Successful classroom deployment of a social document annotation system. In *Proceedings of the sigchi conference on human factors in computing systems*, pages 1883–1892, 2012.