I'm Sure! Automatic Detection of Metacognition in Online Course Discussion Forums

Eddie Huang
Department of Computer Science
University of Illinois
at Urbana–Champaign
Urbana, IL, USA
ezhuang2@illinois.edu

Hannah Valdiviejas
Department of Educational Psychology
University of Illinois
at Urbana–Champaign
Champaign, IL, USA
hsv2@illinois.edu

Nigel Bosch
School of Information Sciences and
Department of Educational Psychology
University of Illinois
at Urbana–Champaign
Champaign, IL, USA
pnb@illinois.edu

***Abstract*—Metacognition is a valuable tool for learning, since it is closely related to self-regulation and awareness of one's own affect. However, methods for automatically detecting and studying metacognition are scarce. Thus, in this paper we describe an algorithm for automatic detection of metacognitive language in writing. We analyzed text from the forums of two online, university-level science courses, which revealed common patterns of phrases that we used for automatic metacognition detection. The algorithm we developed exhibited high accuracy on expert-labeled metacognitive phrases (Spearman's rho = 0.878 and Cohen's kappa = 0.792), and provides a reliable, fast method for automatically annotating text corpora that are too large for manual annotation. We applied this algorithm to analyze relationships between students' metacognitive language and their academic performance, finding small correlations with course grade and medium-sized differences in metacognition across courses. We discuss how our algorithm can be used to advance metacognitive studies and online educational systems.**

***Index Terms*—Metacognition, learning analytics, natural language processing**

I. Introduction

Most of us have thought about our own thinking. We often ask ourselves if an idea is worth pondering over, what to think about next, how we arrived at a flawed conclusion, or if we truly understand a certain idea. While such *metacognition*, or "thinking about thinking", is typically only a tool intended to assist the thinker, humans often behave in ways that communicate signals of their metacognition. In a student-teacher interaction context, for example, teachers may assess metacognition through affective and behavioral cues [1]. Evaluating such expressions of student thinking is also a useful tool for the teacher in this case because they can assess what the student may or may not understand, and adapt their methods of teaching accordingly [2], [3]. In this paper, we present a method for automatic detection of metacognitive language in computerized discussion forums. We apply this method to online STEM (science, technology, engineering, and math) courses, and discuss how it will enable further large-scale research on the interplay between metacognition, affect, and learning.

Metacognition is awareness of one's own thinking [4]. More specifically, it is the process used to plan, monitor and assess one's understanding and thought process [5]. Hence, metacognition is especially important for self-regulated human–computer interaction tasks such as learning with technology [2], [6]. While metacognition is primarily internally-accessible, people may externally indicate their metacognition through bodily gestures (e.g., putting a finger to the chin to indicate thought), conversational cues (e.g., "hmmmmm", "let me think a second") [7], or written language (e.g., descriptions of the thought process in narrative answers to homework exercises) [8]. In educational contexts, teachers can use these indicators of metacognition to understand what students need to work on, which allows teachers to adapt and improve their pedagogy.

Methods for automatic detection are key for enabling large-scale analysis of metacognition in learning contexts (and other discussion-based activities outside the scope of this paper). Detecting metacognition, however, is a nontrivial task. In order to detect metacognition, one must be able to infer some meaning from context. In a parsing algorithm, one of the major difficulties stems from words having different meanings depending on its usage and context. Common research methods employ time-intensive methods in which either the researcher conducts post-activity inquiries to analyze metacognition [9], [10], or participants are required to self-report moments of metacognitive awareness [11]–[15]. Thus, in this paper we focus on *automatic* detection of metacognitive phrases in bodies of student-generated text in ecological contexts (students' daily learning activities). Specifically, we analyze students' metacognition in online STEM courses, though our methods are not limited to these contexts.

Our research is novel in that we provide an algorithm for the automatic detection of metacognition in bodies of text obtained from online discussion forums. By allowing researchers to gather large amount of metacognitive data quickly and accurately, our algorithm can reliably accelerate metacognitive research endeavors. We also discuss later how students and teachers can use automatic detection of metacognition to

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A180211. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

improve learning.

II. Related work

We focus the discussion of related work illustrating the importance of studying metacognition, and examples of related studies that have explored automatic detection of cognitive and affective states related to metacognition.

A. Metacognition, cognition, and affect

Simply put, metacognition is thinking about thinking. It is understanding how much and what information is needed to reach a goal, being conscious of a strategy for obtaining the information required to reach the goal, and thinking about when and how to use the available information to achieve that goal [5]. Although easily confused, cognitive strategies and metacognitive strategies are distinct in an educational context. For example, cognitive strategies include – but are not limited to – rehearsal, elaboration, organization, and critical thinking, while metacognitive strategies deal with time management, self-evaluation, concentration/effort, and self-awareness [16]. While cognitive strategies are critical for making progress and building knowledge, metacognitive strategies oversee these processes and permit improvement to processing (via monitoring), making metacognitive knowledge critical for cognitive strategies to be performed well [4].

A large body of research exists that shows differences between people who possess and/or exercise metacognition and those who do not. Pennycook, Ross, Koehler, & Fugelsang [15] suggested that people who engaged in using metacognition were usually analytic thinkers rather than intuitive thinkers, a framework that is essential to understanding STEM topics. It is metacognitive thinking and strategies that have been shown to improve learning outcomes [17], [18] across the board [19], and have been shown to be the differing variable between high-achievers and low-achievers [6], [20]–[22].

Previous work has also noted that there are key relationships between metacognition and affect. First, self-reports of emotion (which are common in human–computer interaction studies [23]) are correlated with metacognition [24]. Second, metacognitive self-awareness is important for affective self-regulation as well as cognitive self-regulation [25], [26]. These connections between metacognition and affect explain some of the relationships between affect and learning [25], [27], [28], and highlight the importance of further understanding the role of metacognition in current online learning environments.

B. Metacognition and learning

A large body of previous research on metacognition indicates that successful self-regulated learning is conditioned on students' participation in metacognitive activities [2], [11], [29]–[32]. As an educational intervention, the implementation of metacognition has been very successful. Hadie et al. [3] demonstrated that when a traditional lecture was reconstructed so that it contained metacognitive thinking cues, students significantly improved their learning. Casey, Gill, Pennington, & Mireles [33] showed under-performing low-income English Language Learners how to successfully program robots, assisted by metacognitive strategies. Metacognitive elements, like knowing that one does not know something and why they do not know it, are critical in fixing mistakes and dispelling faulty conclusions, especially in educational contexts. Furthermore, metacognition has also been shown to be associated with willingness to adjust faulty learning approaches [34].

C. Automatic detection of metacognition

As previously mentioned, most research on metacognition in the educational context consists of self-reports [10], [14], interventions [8], [35], and observations [1], [36], [37]. While extremely informative, these methods prevent researchers from analyzing spontaneously-produced metacognitive processes on a large scale and in a natural, generalizable setting with no disruptions. To our knowledge, there is no existing research on automatic detection of metacognitive language nor is there research that concerns large-scale analysis of metacognitive language.

Many natural language processing (NLP) methods have been employed to detect related cognitive and affective constructs in student-produced free-form text, however. As previously mentioned, automatic detection from natural language input has potential to be a more accurate, cost-effective, and less time-consuming alternative to self-reports and researcher-annotated responses. Given these advantages, NLP-based automatic detection of cognition and affect is in mainstream usage, including in automatic essay graders like AES Systems [38], Project Essay Grader [39], Latent Semantic Analysis [40], and E-rater [41]. Similarly, NLP-based affect detection has shown promise for evaluating learning [42]. Such methods of automatic detection have proven to be extremely beneficial in the educational context.

The possibility of applying NLP methods to analyze characteristics of learning and learners is also feasible, given their accuracy and relationships to learning. Lintean, Rus, & Azevedo [13] investigated NLP methods on student-articulated paragraphs with hopes of being able to use the paragraphs to describe and better understand students' mental models. They presented several approaches for automatically detecting language to describe students' mental models in an online tutoring environment called MetaTutor. Their study took place during a self-regulatory activity of prior knowledge activation in which the students had to type a paragraph on their prior knowledge of the topic at hand. The results revealed that a word-weighting method using TF-IDF (term frequency–inverse document frequency) features calculated from the text corpus, combined with a Bayes Net machine learning algorithm, yielded the most accurate results.

While researchers have explored automatic detection of affect and cognition in various learning domains including Intelligent Tutoring Systems (ITSs), massive open online courses (MOOCs), and others. Lintean et al.'s work [13] is some of the only work to date (of which we are aware) that automatically analyzes student-produced text dealing with metacognition – though their object was to describe and understand cognition, rather than metacognition, in an educational context designed to promote metacognition. Previous research has also studied automatic detection of constructs related to or including metacognition, such as self-caught mind wandering [12], [43]–[45]. In contrast, our research is novel in that we provide an automatic detection method for metacognitive phrases. Specifically, we capture phrases that indicate *confident* and *unconfident* metacognition, to enable future research into the role of metacognition during text-based interactions with computers.

III. Method

A. Discussion forum post data

Data we analyzed to develop and test the metacognition detection method came from two introductory STEM courses – one with a single semester of data (*course 1*), and another with several semesters of data (*course 2*) – at a large, public university in the Midwestern United States. These data included all of the students' discussion forum posts as well as their final course grades, which were provided to us by university data curators on an coarse-grained ordinal scale to preserve student privacy. Specifically, there were four levels of grades: A- to A+, B- to B+, C- to C+, and D+ or lower. For correlation analyses involving grades, we represented the four grade levels as integers 0 through 3, where higher numbers indicated better grades.

In both courses, forum participation was required as part of students' participation grades. Students were required to regularly post questions they had, or to answer other students' questions. This requirement may have resulted in increased participation (and different quality of participation) relative to online courses with optional forum participation. However, some students did not participate regularly despite this requirement.

B. Participants

We received university ethics board approval before obtaining data analyzed in this paper. Data were retrospectively analyzed (with permission from instructors), to avoid potentially influencing instructor perceptions or students' grades for ongoing courses. University data curators preprocessed all discussion forum posts by replacing student names, places, and other identifying information with placeholders in order to protect their identities. In total, we obtained 19,700 forum posts for analysis, from 710 students who made at least one discussion forum post.

C. Preliminary analysis of metacognition

We began by analyzing a random sample of 200 forum posts to find patterns of metacognitive phrases. We found that most metacognitive phrases began with a pronoun and ended with a thought-related word (e.g., *realized*, *considered*, *thought*). We were interested in detecting metacognitive statements that described students' own metacognitive processes, so we limited pronouns to first person only (e.g., *I*, *we*, *our*). In this initial sample we also noted that metacognitive phrases largely fell into two categories: expressions of a student's thoughts regarding an idea (which we refer to as *confident*) and expressions of doubt about their thoughts (which we refer to as *unconfident*). With few exceptions, metacognitive phrases began with a first-person pronoun and ended with a confident or unconfident thought-related word, and occasionally included negating words in between (e.g, *no*, *not*). We defined "normal" metacognitive phrases as those which followed this pattern.

Our metacognition detection method primarily focuses on capturing this normal pattern of metacognitive phrases, but also captures a select set of common exceptional metacognitive phrases that do not follow the normal pattern.

D. Automatic metacognition detection algorithm

We categorized words into 4 main classes, with some classes containing their own respective subclasses:

1) **Pronouns**
   - *First person*: (e.g., i, we, my)
   - *Non-first person*: (e.g., you, him, she)
2) **Negations**: (e.g., not, no)
3) **Metacognitive**
   - *Confident*: (e.g., understand, know, believe)
   - *Unconfident*: (e.g., confused, unsure, struggle)
4) **Other**: all other words that do not fall into the first three classes

We created a dictionary set of words for each class/subclass to determine whether or not a given word belongs in that category (words that did not appear in any dictionary set we classified as **Other**).

Let w be the list of words for a given body of text (e.g. $w = [i, am, sure, you, are, right]$). We defined the set of **normal metacognitive phrases** within w as

$$\Big\{ w[i, i+j] \quad | \quad w[i] \in \textit{pronouns.first}, \\ w[i+j] \in \textit{metacognitive}, \\ w[k] \notin \textit{pronouns.nonfirst}\ \forall k, i \leq k \leq i+j, \\ \forall i, j : i \leq n, j \leq 5 \Big\}.$$

For example, if $w = [i, am, sure, you, are, right]$, then the set of normal metacognitive phrases would be

$$\textbf{NMP} = \{w[1:3] = [i, am, sure]\}$$

We ensured that phrases within this set did not overlap with each other in the original body of text by left association (we searched for phrases from left to right). We also limited the maximum length of normal metacognitive phrases to five words, as we found this was sufficient to capture virtually all metacognitive phrases of this pattern (see Fig. 1).

Normal metacognitive phrases were further classified into confident and unconfident based on whether the number of negations in the phrase is odd or even, as well as the subclass of the ending metacognitive word. For example, a normal metacognitive phrase with one negation ending with an unconfident metacognitive word would be classified as *confident*

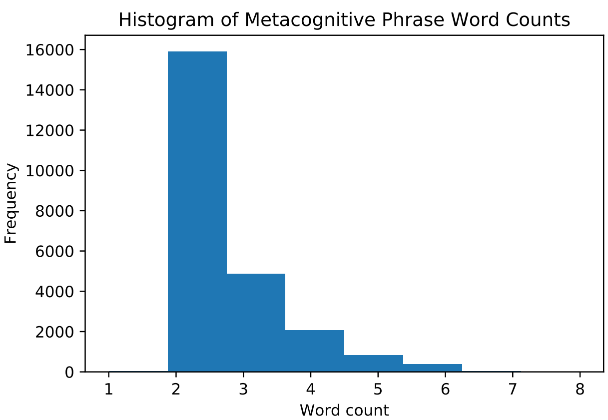

Fig. 1. A histogram of metacognitive phrase lengths (in terms of number of words) as captured by our detection algorithm

and a phrase with three negations that ends with a confident metacognitive word would be classified as *unconfident*.

In order to fully capture all metacognitive phrases, we also created two whitelists of regular expressions, one for confident and another for unconfident, that did not follow the normal pattern – such as acronyms or adages. One example is *IMO*, an abbreviation for "in my opinion", which can indicate metacognitive awareness.

The largest difficulty in developing the algorithm was curating the dictionary word lists that altered the behavior of the algorithm. Words like "lost" were often too ambiguous to be considered metacognitive, and so we chose to capture common phrases containing those words with the regular expression whitelists. In total, we analyzed 1000 posts to curate the dictionary word lists and assess the algorithm. Fig. 2 and Fig. 3 illustrate the results of applying the algorithm to two randomly-chosen forum posts.

`[REDACTED], Although your estimate doesn't agree with mine of 20, I believe you bring up a very valid point. I had not even considered that the star formation rate was much faster long ago than it is now. Many of the stars in the galaxy that we can see today could very well have been born billions of years ago. It's a smart idea to consider not just rely on what we can see in the sky, but on the changes in the galaxy that occur. Although with this new information I do not know what my new estimate would be, but I believe that there is undoubtedly merit in your reasoning`

Fig. 2. Algorithm's annotation of metacognitive language from one example forum post, with confident phrases highlighted in green and unconfident phrases in red.

`Excellent job [REDACTED]! As we have had to do in the past for estimates, we tend to rely on what we know from Earth for it is our only known example. Seeing as we have just started to scratch the surface of communication I believe we will be communicating for many more years. My estimate may be larger than yours but we both had the same thought process deciding our estimate of L!`

Fig. 3. Another example of automatic annotation on a forum post with only confident phrases (highlighted in green).

E. Algorithm assessment

As stated before, we began by randomly sampling 200 forum posts. We had two researchers who were familiar with metacognition research literature – but not with the algorithm – evaluate and label the true number of confident and unconfident phrases for 100 of these posts (half of the sample). The annotators resolved their disagreements, and we utilized their annotations to develop the first iteration of the automatic labeling algorithm. The annotators then labeled the second half of the sample, which we used as the initial test of the prototype algorithm.

Within this second sub-sample of 100 posts, the algorithm achieved a Cohen's κ of 0.770 with a 95% confidence interval (CI) of $[0.684, 0.855]$, and Spearman's ρ correlation of 0.871 (CI $=$ $[0.815, 0.912]$) for positive metacognitive phrases [46], [47]. The algorithm achieved Cohen's $\kappa = 0.804$ (CI $=$ $[0.665, 0.943]$) and Spearman's $\rho = 0.817$ (CI $=$ $[0.740, 0.873]$) for negative metacognitive phrases.

However, the confidence intervals were relatively large (exceeding ± 0.100 for kappa, for example). By scaling our sample size through repetition (Fig. 4 and Fig. 5), we extrapolated that another test of 400 posts were needed to assess the algorithm's quality with a sufficiently tighter confidence interval (well under ± 0.100 in all cases). Thus, the annotators labeled another 400 randomly-chosen posts, and we made algorithmic refinements based on the second half of the 200-post sample.

From this second round of testing with 400 posts, we indeed achieved a much tighter confidence interval with κ and ρ values that indicate large effect sizes (at least 0.800 in all respects) [48].

During the second round of testing, we found it necessary to modify the lists of words and phrases – predominately for unconfident metacognitive phrases. Given that we modified the algorithm, we completed another round of manual annotation and testing following the same procedure, again with 400 forum posts. This last sample of 400 posts represents the test set, which we analyzed for our final results.

IV. Results

The focus of this paper is on automatic detection of metacognition. However, in these results we also consider

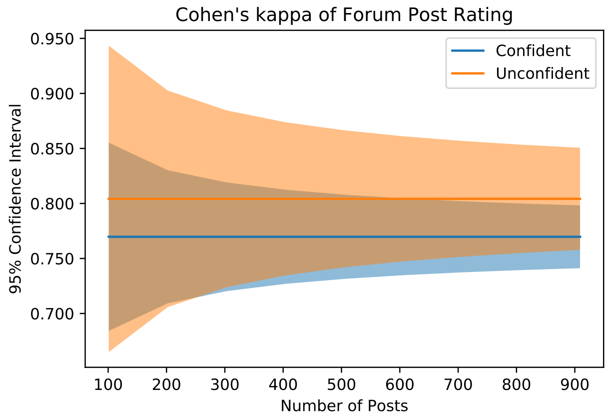

Fig. 4. An extrapolation of the κ confidence interval from an initial test on 100 sample discussion forum posts.

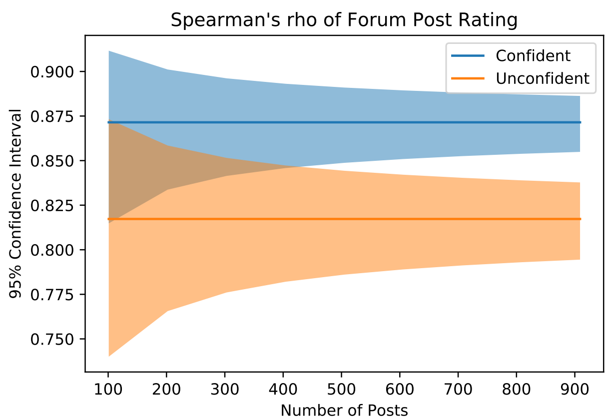

Fig. 5. An extrapolation of the ρ confidence interval from an initial test on 100 sample discussion forum posts.

example applications of the metacognition detection method to demonstrate its utility as a measure.

A. *Metacognition detection accuracy*

Our metacognitive language detection algorithm demonstrated high accuracy in predicting the number of confident and unconfident metacognitive phrases in bodies of text. We compared the finalized algorithm to two human annotators on a held-out test set of 400 forum posts from an online STEM course (course 1). Each human rater independently annotated the posts and resolved any disagreements via discussion with a third rater to create a consensus on the ground truth annotations. Finally, we calculated the agreement between the two human raters (before disagreement resolution) as well as the agreement between the algorithm's output and the resolved ground truth annotations.

We found that our algorithm's accuracy versus the ground truth annotations was comparable to the agreement between the two human raters (Table I), who were themselves highly reliable [48]. Finally, when applied to the entire dataset from both courses (19,700 forum posts), the algorithm detected 23,466 metacognitive phrases (20,092 confident and 3374 unconfident).

TABLE I
SPEARMAN'S RHO AND COHEN'S KAPPA AGREEMENT BETWEEN OUR ALGORITHM AND HUMAN ANNOTATORS (AFTER CONSENSUS), AS WELL AS BETWEEN HUMAN RATERS PRE-CONSENSUS ON THE NUMBER OF METACOGNITIVE PHRASES PER FORUM POST (ALL $p < .001$).

	Algorithm accuracy		**Human agreement**	
Metacognition type	ρ	κ	ρ	κ
Confident	.861	.776	.893	.810
Unconfident	.859	.825	.786	.772
Either	.878	.792	.899	.822

B. *Correlations with learning outcomes*

We applied the detection algorithm to all data from both courses and examined correlations between metacognitive language usage and final course grade (a measure of success). In particular, we considered the mean number of metacognitive phrases per forum post for each student, as well as means for confident and unconfident phrases separately. Grade levels were ordinal variables, so we computed rank-order correlations (Spearman's ρ) for these comparisons.

Results in Table II show that there were small correlations between metacognition and grade within each of these courses. In course 1, confident metacognitive phrases and overall metacognitive phrase usage were positively correlated with grade, indicating higher grades for students who used more metacognitive language. In course 2, unconfident metacognitive phrases were positively related to grade, though each of these effect sizes was small.

Conversely, effect sizes for the relationship between metacognition and grade were medium-sized [48]. In the next section, we explore this notable difference in more detail.

TABLE II
SPEARMAN CORRELATIONS BETWEEN METACOGNITIVE PHRASE OCCURRENCES (PER POST) AND FINAL GRADES. * INDICATES SIGNIFICANT CORRELATIONS ($p < .05$).

Course	**Confident** ρ	**Unconfident** ρ	**Either** ρ
Course 1	.158*	.128	.166*
Course 2	.077	.090*	.071
Combined	.473*	.358*	.468*

C. *Metacognitive differences across courses*

Correlations between metacognition and grade were surprisingly large when combining data from both courses (ρ = .358, .473, and .468 for confident, unconfident, and total metacognition, respectively). On the other hand, within-course correlations (i.e., controlling for course) were comparatively low (Table II). We investigated this pattern further and found that the contrasting correlations were attributable to differences between the two courses.

As shown in Table III, students in course 1 employed significantly more metacognitive phrases than students in course 2. Furthermore, the course 1 students received significantly higher grades than course 2 students, thereby explaining the substantial correlations between metacognition and grade when analyzing both courses combined[1].

TABLE III
METACOGNITION AND GRADE DIFFERENCES ACROSS COURSES. GRADES ARE ON A $[0, 3]$ SCALE, WHERE HIGHER NUMBERS ON THE SCALE INDICATE BETTER GRADES. ALL DIFFERENCES BETWEEN COURSES WERE STATISTICALLY SIGNIFICANT ($p < .001$).

Measure	**Course 1 mean**	**Course 2 mean**
Metacognitive phrases per post		
Confident	1.393	0.205
Unconfident	0.203	0.092
Either	1.596	0.297
Grade	2.662	1.381

V. DISCUSSION

A. Main findings

We were interested in the problem of automatic detection and annotation of metacognitive language in computerized educational discussion forums. To approach this problem, we examined human-annotated forum text and developed a natural language processing approach that locates metacognitive phrases in bodies of text with high accuracy.

We also measured correlations between metacognitive language and academic performance. While we expected a positive relationship, we found relatively small correlations between metacognitive language and academic performance within each course. On the other hand, when the courses were combined, correlations between metacognition and grades were much larger, which suggests important differences between courses. Possible explanations for this could include the reasons students had for taking a particular course, course structure/curriculum, student demographic differences, influence of forum participation on grade, etc. These course differences will be the subject of our further research.

Our algorithm can also be used in future metacognitive research to automate text analysis of metacognitive language. When there is a relatively large amount of data, humans may incur error from fatigue, inexperience, or other causes – whereas our algorithm is inherently consistent, efficient, and validated on a large sample of reliable expert annotations.

[1]Differences in metacognition between courses remain similar after controlling for length of forum posts; thus, correlations between metacognition and grade remain as well.

B. Limitations and future work

While our metacognition detection method exhibited high agreement with human raters, its ability to generalize to notably different sources of text remains unclear. For example, individuals who are much older or younger than typical university students may express metacognition differently. Future work is needed to measure generalization at different levels including across universities, ages, demographic characteristics, and other dimensions. Additionally, our data exclusively consisted of English-language text, and our current detection model would certainly not generalize to other languages.

Our algorithm looks for metacognitive language by finding specific patterns of particular words. It is thus primarily limited by the words in the word lists. Therefore, in future work we plan to address issues of generalization, as they arise, by modifying these lists that help the algorithm categorize metacognitive phrases.

Another alternative solution is deep learning, a popular method with many successful applications. While our solution already achieves high correlations, it might be worth training a neural network to compare against our solution.

In future work we also intend to apply the algorithm to a variety of online STEM courses, with the goal of identifying metacognitive differences between students and courses that might lead to educational improvements. For example, interventions intended to improve learning outcomes via metacognitive interventions might likely be much more valuable for students who are not already evaluating their own knowledge and thought processes.

C. Concluding remarks

Metacognitive research is limited by the amount of data we can gather. Current research relies on human annotation, which can be time consuming and inconsistent across research endeavors. Our algorithm consistently detects metacognition from text generated during computer-mediated discussion, and has shown high accuracy in a large, held-out test set (400 posts). This method will enable future research to quickly measure metacognitive language usage, thereby enhancing our understanding of the relationships between metacognition, affect, and learning. Toward this goal, we have made the source code of our algorithm publicly available[2], as well as a web interface to enable users to easily apply the algorithm to their own data files.

REFERENCES

[1] M. Goos, "Understanding metacognitive failure," *The Journal of Mathematical Behavior*, vol. 21, no. 3, pp. 283–302, Jan. 2002.

[2] A. Zohar and S. Barzilai, "A review of research on metacognition in science education: Current and future directions," *Studies in Science Education*, vol. 49, no. 2, pp. 121–169, Sep. 2013.

[3] S. N. H. Hadie, A. Hassan, Z. I. M. Ismail, H. N. Ismail, S. B. Talip, and A. F. A. Rahim, "Empowering students' minds through a cognitive load theory-based lecture model: A metacognitive approach," *Innovations in Education and Teaching International*, vol. 55, no. 4, pp. 398–407, Jul. 2018.

[2]Source code: https://github.com/aigagror/metacognitive_phrase_detector
User interface: https://www.aigagror.com/research/metacognition/

[4] J. H. Flavell, “Metacognition and cognitive monitoring: A new area of cognitivedevelopmental inquiry,” *American Psychologist*, vol. 34, no. 10, pp. 906–911, 1979.

[5] I. Brinck and R. Liljenfors, “The developmental origin of metacognition,” *Infant and Child Development*, vol. 22, no. 1, pp. 85–101, 2013.

[6] S. W. Vanderstoep, P. R. Pintrich, and A. Fagerlin, “Disciplinary differences in self-regulated learning in college students,” *Contemporary Educational Psychology*, vol. 21, no. 4, pp. 345–362, Oct. 1996.

[7] L. W. Barsalou, “Grounded cognition,” *Annual Review of Psychology*, vol. 59, no. 1, pp. 617–645, 2008.

[8] B. L. Casselman and C. H. Atwood, “Improving general chemistry course performance through online homework-based metacognitive training,” *Journal of Chemical Education*, vol. 94, no. 12, pp. 1811–1821, Dec. 2017.

[9] J. A. Cardinale and B. C. Johnson, “Metacognition modules: A scaffolded series of online assignments designed to improve students’ study skills,” *Journal of Microbiology & Biology Education*, vol. 18, no. 1, Apr. 2017.

[10] K. S. McCarthy, A. D. Likens, A. M. Johnson, T. A. Guerrero, and D. S. McNamara, “Metacognitive overload!: Positive and negative effects of metacognitive prompts in an intelligent tutoring system,” *International Journal of Artificial Intelligence in Education*, vol. 28, no. 3, pp. 420–438, Sep. 2018.

[11] A. Vrugt and F. J. Oort, “Metacognition, achievement goals, study strategies and academic achievement: Pathways to achievement,” *Metacognition and Learning*, vol. 3, no. 2, pp. 123–146, Aug. 2008.

[12] R. Bixler, N. Blanchard, L. Garrison, and S. K. D’Mello, “Automatic detection of mind wandering during reading using gaze and physiology,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, ser. ICMI ’15. New York, NY: ACM, 2015, pp. 299–306.

[13] M. Lintean, V. Rus, and R. Azevedo, “Automatic detection of student mental models based on natural language student input during metacognitive skill training,” *International Journal of Artificial Intelligence in Education*, vol. 21, no. 3, pp. 169–190, Jan. 2012.

[14] M. M. Spada and G. Caselli, “The Metacognitions about Online Gaming Scale: Development and psychometric properties,” *Addictive Behaviors*, vol. 64, pp. 281–286, Jan. 2017.

[15] G. Pennycook, R. M. Ross, D. J. Koehler, and J. A. Fugelsang, “Dunning-Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence,” *Psychonomic Bulletin & Review*, vol. 24, no. 6, pp. 1774–1784, Dec. 2017.

[16] S. Kuama and U. Intharaksa, “Is online learning suitable for all English language students?” *PASAA: Journal of Language Teaching and Learning in Thailand*, vol. 52, pp. 53–82, 2016.

[17] J. L. Nietfeld, L. Cao, and J. W. Osborne, “Metacognitive Monitoring Accuracy and Student Performance in the Postsecondary Classroom,” *The Journal of Experimental Education*, vol. 74, no. 1, pp. 7–28, 2005.

[18] R. Azevedo and J. G. Cromley, “Does training on self-regulated learning facilitate students’ learning with hypermedia?” *Journal of Educational Psychology*, vol. 96, no. 3, pp. 523–535, 2004.

[19] M. Cardelle-Elawar, “Effects of metacognitive instruction on low achievers in mathematics problems,” *Teaching and Teacher Education*, vol. 11, no. 1, pp. 81–95, Jan. 1995.

[20] A. Bandura, *Self-efficacy: The Exercise of Control*. New York, NY: W. H. Freeman, 1997.

[21] B. J. Zimmerman, “Becoming a self-regulated learner: An overview,” *Theory Into Practice*, vol. 41, no. 2, pp. 64–70, 2002.

[22] G. Horowitz, L. A. Rabin, and D. L. Brodale, “Improving student performance in organic chemistry: Help seeking behaviors and prior chemistry aptitude,” *Journal of the Scholarship of Teaching and Learning*, pp. 120–133, Aug. 2013.

[23] R. A. Calvo and S. K. D’Mello, “Affect detection: An interdisciplinary review of models, methods, and their applications,” *IEEE Transactions on Affective Computing*, vol. 1, no. 1, pp. 18–37, Jan. 2010.

[24] M. M. Spada, A. V. Nikevi, G. B. Moneta, and A. Wells, “Metacognition, perceived stress, and negative emotion,” *Personality and Individual Differences*, vol. 44, no. 5, pp. 1172–1181, Apr. 2008.

[25] A. Efklides, “Metacognition and affect: What can metacognitive experiences tell us about the learning process?” *Educational Research Review*, vol. 1, no. 1, pp. 3–14, Jan. 2006.

[26] ——, “Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL model,” *Educational Psychologist*, vol. 46, no. 1, pp. 6–25, Jan. 2011.

[27] N. Keith and M. Frese, “Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects,” *Journal of Applied Psychology*, vol. 90, no. 4, pp. 677–691, 2005.

[28] M. Malekzadeh, M. B. Mustafa, and A. Lahsasna, “A review of emotion regulation in intelligent tutoring systems,” *Educational Technology & Society*, vol. 18, no. 4, pp. 435–445, 2015.

[29] P. Georghiades, “Beyond conceptual change learning in science education: Focusing on transfer, durability and metacognition,” *Educational Research*, vol. 42, no. 2, pp. 119–139, Jan. 2000.

[30] S. Hallam, “The development of metacognition in musicians: Implications for education,” *British Journal of Music Education*, vol. 18, no. 1, pp. 27–39, Mar. 2001.

[31] W. Schneider and C. Artelt, “Metacognition and mathematics education,” *ZDM*, vol. 42, no. 2, pp. 149–161, Apr. 2010.

[32] G. Schraw, K. J. Crippen, and K. Hartley, “Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning,” *Research in Science Education*, vol. 36, no. 1, pp. 111–139, Mar. 2006.

[33] J. E. Casey, P. Gill, L. Pennington, and S. V. Mireles, “Lines, roamers, and squares: Oh my! using floor robots to enhance Hispanic students understanding of programming,” *Education and Information Technologies*, vol. 23, no. 4, pp. 1531–1546, Jul. 2018.

[34] J. D. Stanton, X. N. Neider, I. J. Gallegos, and N. C. Clark, “Differences in metacognitive regulation in introductory biology students: When prompts are not enough,” *CBELife Sciences Education*, vol. 14, no. 2, pp. ar15:1–ar15:12, Jun. 2015.

[35] R. Visser and A. B. Flynn, “What are students’ learning and experiences in an online learning tool designed for cognitive and metacognitive skill development?” *Collected Essays on Learning and Teaching*, vol. 11, pp. 129–140, 2018.

[36] A. F. Gourgey, “Metacognition in basic skills instruction,” *Instructional Science*, vol. 26, no. 1, pp. 81–96, Mar. 1998.

[37] R. B. Thompson and B. J. Foster, “Socioeconomic status and parentchild relationships predict metacognitive questions to preschoolers,” *Journal of Psycholinguistic Research*, vol. 43, no. 4, pp. 315–333, Aug. 2014.

[38] M. D. Shermis and F. D. Barrera, “Exit assessments: Evaluating writing ability through automated essay scoring,” Florida International University, Tech. Rep., 2002.

[39] L. M. Rudner and T. Liang, “Automated essay scoring using Bayes’ theorem,” *The Journal of Technology, Learning and Assessment*, vol. 1, no. 2, Jun. 2002.

[40] T. K. Landauer, D. S. McNamara, S. Dennis, and W. Kintsch, *Handbook of Latent Semantic Analysis*. New York, NY: Routledge/Taylor and Francis, 2011.

[41] J. Burstein, “The E-rater scoring engine: Automated essay scoring with natural language processing,” in *Automated Essay Scoring: A Cross-disciplinary Perspective*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers, 2003, pp. 113–121.

[42] S. Crossley, L. Paquette, M. Dascalu, D. S. McNamara, and R. S. Baker, “Combining click-stream data with NLP tools to better understand MOOC completion,” in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, ser. LAK ’16. New York, NY, USA: ACM, 2016, pp. 6–14.

[43] M. Faber, R. Bixler, and S. K. D’Mello, “An automated behavioral measure of mind wandering during computerized reading,” *Behavior Research Methods*, vol. 50, no. 1, pp. 134–150, 2018.

[44] N. Bosch and S. K. D’Mello, “Automatic detection of mind wandering from video in the lab and in the classroom,” *IEEE Transactions on Affective Computing*, in press.

[45] P. Pham and J. Wang, “AttentiveLearner: Improving mobile MOOC learning via implicit heart rate tracking,” in *Artificial Intelligence in Education*, ser. Lecture Notes in Computer Science. Cham, CH: Springer, Jun. 2015, pp. 367–376.

[46] J. Cohen, “A coefficient of agreement for nominal scales,” *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960.

[47] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by-Step Approach*, 2nd ed. Hoboken, NJ: John Wiley & Sons, Apr. 2014.

[48] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum, 1988.