Generating Learning Analytics to Improve Learners’ Metacognitive Skills Using nStudy Trace Data and the ICAP Framework

Zahia Marzouk zmarzouk@sfu.ca
Mladen Rakovic mrakovic@sfu.ca
Philip Winne winne@sfu.ca
Simon Fraser University 8888 University Drive, Burnaby, V5A 1S6 Canada
Telephone: 001-778-782-4548

ABSTRACT

Metacognition is integral in the cycle of self-regulated learning. Enhancing learner’s metacognitive skills is a focus of many studies. Offering learning analytics to learners about learning has also gained popularity as a means to improve their metacognitive skills. A key question is, What types of analytics can meaningfully prompt change? We propose Chi’s (2009; 2014) ICAP framework can lend meaning to analytics and guide learners to enhance metacognitive skills.

CCS Concepts: Software notations and tools

Keywords: Metacognition, learning analytics, social learning networks, traces, writing analytics

1. INTRODUCTION

Metacognition is cognition about one’s cognitive processes [2] and mental states (knowledge, feelings, and other thoughts), learners’ “awareness of their own cognitive machinery and how the machinery works” [21 p 5]. A vibrant research topic [14], it accounts for nearly 17% of variance in learning among learners of different ages and backgrounds across various tasks. In contrast, intellectual abilities account for 10% [32].

Metacognitive monitoring involves observing learning activities, strategy use and performance; and judging the fit between goals and a current state. It is critical in self regulated learning (SRL) as it sets the stage for metacognitive control that adapts strategies, affect, and behavior [37]. A self-regulating learner engaging in metacognitive monitoring and control is an empowered learner. SRL starts with awareness. Learners can not adapt if they are not cognizant of their current status. When learners are aware of their learning activities, the probability of change increases [34]. Learners who lack a full and accurate record of the frequency, intensity or quality of behavior or performance must rely on selective and imperfect memories leading to mistaken perceptions of how they study [28]. This has elevated interest in recording learners’ behaviors as they study then providing analytics about their learning.

Analytics provide information learners can use to effectively monitor and control learning. Thus enhancing metacognitive skills is potentially a key to productive SRL [20]. Although use of analytics is widespread in business, marketing and scientific research, it is sporadic in education [11].

One well-established application of analytics in education is the Course Signals program at Purdue University [1]. This program uses trace data collected by Purdue’s learning management system (LMS) together with data from the student information system (SIS) to identify students at risk of failing courses. This program’s success rests on academic analytics [1] rather than learning analytics. Academic analytics use students’ data from a LMS (e.g., frequency of log in, contributions to class discussions) together with statistical techniques and prediction models to inform decisions [7] but they lack data about the actual process of learning. Gasevic et al.’s (2015) article “Let’s not forget: Learning analytics are about learning” emphasizes this distinctive feature of learning analytics. Learning analytics support processes learners engage to learn.

The foundation of any analytics is data [7]. What data are needed? Pistilli, Willis and Campbell (2014, p. 85) answer “meaningful, useful and obtainable data” that meet principles set out by Chickering and Gamson (1987):

1. Encourage contact between students and faculty
2. Develop reciprocity and cooperation among students
3. Encourage active learning
4. Give prompt feedback
5. Emphasize time on task
6. Communicate high expectations
7. Respect diverse talents and ways of learning

Principle 3, “encourage active learning,” reflects a student-centered approach [22]. Active learning is “anything course related that all students in a class session are called upon to do other than simply watching, listening and taking notes” [15, p. 2]. Pistilli et al. (2014) suggest creating a learning environment in which learners interact with content, and are prompted and guided to reflect on learning processes and products. These are optimal conditions for eliciting and collecting meaningful data.

We propose: (1) using the nStudy learning system to unobtrusively collect trace data as learners interact with content in meaningful activities, (2) adopting the ICAP (Interactive-Constructive-Active-Passive) framework of active learning to categorize data [8,9], and (3) providing analytics computed from nStudy’s trace data to support metacognition.

nStudy

How data are collected determines their quality [24]. Pistilli et al. (2014) noted that gathering self-report data from learners alerts them to data collection. Unlike most data a LMS collects, self-reports intrude in the learning process. Data collected by an LMS typically include students’ grades, log-in events, downloads, and participation in online discussions. These data are “ambient,” i.e., collected as a learner participates naturally in course activities. Ambient data gathered are unobtrusive [19] and ubiquitous [12]. Such data are used to generate analytics in Purdue’s Signals program [1].

1 http://www.sfu.ca/edpsychlab/nstudy.html

Copyright © 2016 for the individual papers by the papers’ authors. Copying permitted only for private and academic purposes. This volume is published and copyrighted by its editors. LAL 2016 workshop at LAK ’16, April 26, 2016, Edinburgh, Scotland.
LMS data are common and accessible but are too coarse to provide insights into processes students use in learning. They do not describe specifically what students did or might do differently when they study in the future [26]. Gasevic, Mirriati, Dawson and Joksimovic (2014) reported that counting the frequency of studying operations a learner performs while using a video annotation tool falls short of a sufficient measure of the quality of learning products. Time on task and raw frequency of studying activities insufficiently capture key qualities of learning [17].

Data gathered by an LMS rarely reveal strategies learners use; when and how they search for information and what information they search for; when they monitor learning, etc. To fill these gaps, we developed nStudy, an online learning system in which learners highlight text, create notes, tag, organize and search for saved information in everyday learning. Learners can re-use nStudy’s artifacts in drafting essays. nStudy facilitates sharing information and co-constructing knowledge by a chat/discussion feature, the hub.

As learners work, nStudy unobtrusively collects ambient traces – time stamped, very fine-grained data about operations learners apply (e.g., highlighting, tagging, note-taking) and information operated on (e.g., text highlighted, tags applied, content contributed to a discussion). Ambient, meaningful, and nonintrusive data are the kind of data needed to generate learning analytics that enhance self-regulated learning (SRL).

**The ICAP Framework**

How should nStudy’s data be interpreted? Brooks, Greer and Gutwin (2014) noted the importance of meaningfully labeling data. The ICAP framework that describes engagements in active learning meets this criterion. Pardo (2014) identified three levels of engagement: behavioral, emotional, and cognitive. Here, we focus on cognitive engagement, “the amount and type of strategies that learners employ” [33, p. 4] and how learners are strategic and self-regulating [35][36].

The ICAP framework [8, 9] posits four modes of cognitive engagement: passive, active, constructive and interactive. Each mode is identified by learners’ overt behaviors as they study. ICAP also relates underlying cognitive processes to each mode and, on that basis, predicts different levels of learning. We use the ICAP framework because it clearly describes learners’ overt learning behaviors and cognitive processes associated with these behaviors, and it is well supported by research.

The passive mode is defined as “learners receiving information without overtly doing anything related to learning” [9, p. 221], e.g., listening to a lecture without taking notes. “Attending” cognitive processes that underlie overt behaviors include storing information episodically without integrating it with prior knowledge or classifying it using schemas. The active mode implies learners do something with their hands or bodies when learning, for instance, copying definitions or highlighting text. Possible “gap filling” covert cognitive processes involve activating prior knowledge and assimilating new knowledge into pre-existing schemas. The constructive mode is when learners actively create meaning by generating information beyond what was presented or initially known, e.g., when learners draw a concept map. Underlying cognitive processes are “generating” processes [9, p. 228] that integrate new information with prior knowledge, “elaborating each other’s contributions, incorporate feedback and perspectives, challenge & requesting explanations, resolving conflicts” [p. 13]. When the learner exchanges information with peers or a learning system, the interactive mode is activated provided this “outside” information is used to construct knowledge. The underlying cognitive processes are “mutually generative” [9] and involve incorporating feedback and considering new ideas.

Under ICAP, passive engagement results in “minimal understanding,” active engagement in “shallow understanding,” constructive engagement in “deeper understanding that might transfer” and interactive engagement in “understanding that might innovate novel ideas” [p. 14]. This is a straightforward ordinal classification: interactive > constructive > active > passive.

## 2. SUGGESTED LEARNING ANALYTICS

### 2.1. Analytics about Studying in General

Note that learning analytics proposed here are theoretically grounded but await empirical testing.

Learners’ operations in nStudy articulate to ICAP. For example, the passive mode is indicated when a learner accesses URLs but does not operate on content (e.g. highlight, tag, create notes). The active mode is indicated when a learner creates a note by copying and pasting content from a source. The constructive mode is indicated when content in a note is original. The interactive mode is indicated when a learner constructively exchanges information with peers in the hub.

**Part 1: General Study View (Fig. 1):**

Analytics will include: (a) A recommendation that has support in research regarding each mode of engagement, (b) a pie chart presenting a classification of the learner’s study operations in each of ICAP’s modes, and (c) metacognitive prompts to help learners reflect on their studying.

![Figure 1. General Study View](image)

**Part 2: Specific Study View:**

For a more detailed report of study activities within each mode, a learner can press the “Specific Study View” button (Fig. 1).
A menu shows four options: Passive, Active, Constructive and Interactive. If the learner chooses “Passive,” the percent of studying operations in this mode shows as a bar chart (Fig. 2)

**Theoretical Foundation for Analytics:**

*a. Provide Meaningful Feedback:* Pistilli, Willis and Campbell (2014) posited feedback to learners needs to be meaningful and actionable. Brooks, Greer and Gutwin (2014, pp. 124) noted the importance of providing an “individualized learning experience.” Ipsative (within person across time) feedback is individualized but what kinds of ipsative feedback can benefit learning? Imagine providing this analytic to a learner: “Today you created 40 quotes, 3 summary notes, accessed 12 URLs, while three days ago you made 3 contributions to the hub and created 2 notes.” Does this provide meaningful information about learning? How would this help her to metacognitively monitor her studying behavior? The ICAP framework offers a solution where each mode of cognitive engagement serves as a meaningful description of learning operations expressed in terms of nStudy operations, as in Fig. 1.

*b. Enhance Learners’ Metacognitive Monitoring:* Long and Siemens (2011) view the real value of learning analytics as guiding decisions about learning; taking action is integral to learning analytics. Wise, Zaho and Hausknecht (2013) suggested showing learners reports about their participation in online class discussions and helping them reflect on it. Then it is up to learners to choose what action to take. We propose showing learners analytics describing operations engaged during learning and prompting them to consider their learning behavior without providing recommendations about what they should do. This gives learners opportunity to consciously assess, reflect on and decide what to do next, i.e., to self regulate learning via metacognitive monitoring and metacognitive control. Some research reports that consciously attending to and purposefully assessing one’s behavior is an effective behavior management technique [27]. Lan, Brady and Par (1993) reported that self-monitoring students performed better than students monitored by their instructors. A meta-analysis by Weber and colleagues (1993) found that special education students who engaged in self-monitoring behavior made more changes to behavior, which led to better performance, than students who did not. Although monitoring is important to learning, students are not very efficient in monitoring learning on their own [e.g. 6]. Instead of telling learners what action to take, we provide tools to help them manage their learning such as (a) cognitive prompts to guide their studying (e.g., prompts in the summary note template presented later) and (b) metacognitive prompts to scaffold their monitoring and control of information processing [3]. The two questions provided in Fig 1 prompt the learner to reflect on studying, clarifying their current status.

2. 2. Analytics for Summary Notes:

Schunk (1985) pointed out that training learners in different learning strategies then giving them opportunity to choose what they judge to be the most effective one enhances learners’ understanding of the task. According to the study strategies literature, writing summaries promotes learning [14]. Summarizing involves constructive engagement as learners read, identify, rephrase and synthesize important information [8]. However, the effect of writing a summary on learning is tied to the summary’s quality. Summaries that omit important information or include wrong information do not promote learning [4]. Summaries benefit learning when they include all main ideas in a text and link main ideas to prior knowledge [14]. This mirrors the generative model of learning – when learners relate information they study to prior knowledge, information becomes meaningful and more memorable [38].

Building on these findings, we propose an nStudy summary note template in Fig. 3 with prompts that scaffold learners’ constructive information processing to enter information into two fields. The first prompt asks the learner to provide a main idea; the second prompt requests an elaboration. Trevors, Duffy and Azevedo (2014) report that prompted notes benefit learners more than non-prompted ones.

![Figure 3. A summary note template](image)

(b) Analytics about Summary Notes:

We propose analytics that inform learners how they have created summary notes. According to the ICAP framework and in the context of nStudy, learners’ notes could be classified as “Active” or “Constructive.” Table 1 operationally defines how we classify learners’ trace data about summary notes. Fig. 4 shows analytics for learners’ general usage of summary notes.

<table>
<thead>
<tr>
<th>Field</th>
<th>Learner’s trace data of:</th>
<th>Active</th>
<th>Constructive</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Main Idea</td>
<td>Copied</td>
<td>Original</td>
</tr>
<tr>
<td>2</td>
<td>How does it relate ... ?</td>
<td>Empty</td>
<td>Completed</td>
</tr>
</tbody>
</table>

![Figure 4. Analytics about summary notes](image)

II. Analytics of trace data provided in “Field 1: Main Idea” (Fig. 5)

First, instructors identify main ideas in the text using nStudy’s targets feature. (A target is text an instructor tags as belonging to class. nStudy logs operations on targets.) To make a summary note, the learner selects some text and chooses the summary note template. nStudy compares the learner’s rendering of “main idea” to the selected and adjacent text in the source to verify whether text was copied or paraphrased.

![Figure 5. Analytics for providing main ideas](image)
III. Analytics of trace data of “Field 2: How does it relate to what you already know?”
nStudy records a learner’s response to the prompt relating the main idea to prior knowledge, then provides analytics in Fig. 6.

The diagram shows claims, evidence and examples the learner created. Empty boxes mark gaps. For instance, according to Fig. 8, the learner needs to find evidence and examples for claim 2. The diagram uses yellow and blue colors to display parts the learner used and did not use in the essay created with nStudy. The metacognitive prompt “Is there anything you think you need to add?” encourages reflecting on artifacts and the learning product (the essay) and to decide whether it needs elaborating.

2.4. Social Learning Network Analysis:
Among principles of connectivistic theory, Siemens (2004) lists diversity of opinions and connecting of specialized sources of information. Therefore, learners should be able to access, adopt or critically evaluate specific knowledge offered by other learners in their social network. However, as a network grows in complexity, it can be challenging for learners to use its social capital [25], so it is important to filter all generated messages to show only those that make a real contribution to the exchange rather than count all messages sent and received. This simplifies the network and facilitates access to knowledge. We use the interactive part of ICAP framework as our reference point for creating a social learning network that can offer learners analytics about: a) their interactive contributions to the learning hub with regard to any topic and b) a topic-specific view of exchanges in the hub. The learning analytic displayed a social learning network as an undirected graph of nodes and weighted edges. Larger nodes represent learners with a greater number of contributions. nStudy uses node centrality and betweenness centrality as metrics for contributions.

In nStudy’s hub, a student can create a new discussion topic, discuss an existing topic or reply to other students’ posts. Prior to adding a comment under a particular discussion in nStudy’s learning hub, Jane is provided with menu to select the type of comment she wants to post: Disagree, Give Reason, Request Justification, Ask a Question, Elaborate or Share (Fig. 9).

After selecting her tag, Jane writes a comment providing reasons that support her contribution. For example, in a discussion about linear regression, she selected “Share” and commented “At this link you can watch a video explaining interaction.” nStudy dynamically tracks discussions in the hub and applies text mining algorithms to identify relevant posts corresponding to a particular topic. For instance, Jane’s post is relevant to the topic of linear regression. nStudy excludes irrelevant posts from a learning analytics report. When Nelson replies to Jane’s post tagging it “Elaborate” then posts “Thanks,” this is classified as an irrelevant post. His message to Jane is not included in computations that create learning analytics about the interactive mode of ICAP. In contrast, Lee
Nelson, among the most influential participants under this topic, metacognitively analysed his knowledge gap and commented: “And what happens if variables are transformed?” wrongly tagging this post as “Disagree”. However, using text mining algorithms nStudy will consider Lee’s post as relevant and include it in the social learning network and categorize his post as “Ask a Question” in the software’s database.

Fig. 10 shows a graph in which learners are represented as nodes. Edges are created according to two types of responses: response to a discussion topic’s question and responses to other student’s post (reply). For instance, Jane directly responded to the discussion topic. As every topic has its creator, nStudy considers Jane’s message as a message to the topic’s creator and the edge between these two participants in the graph is weighted as 1 [16]. When Jane submits her next relevant post to the creator or the creator responds directly to Jane, the weight of the edge connecting their nodes increases by 1. In addition, this concept limits the role of a topic’s creator – she can not respond directly to topic s/he created, but can reply to learners’ posts.

Two types of social learning networks are generated by nStudy. The network in Fig. 10 represents a general discussion across all topics in the hub. Using this graph, the learner and the course instructor can easily identify engaged students and gain better insight to network structure, such as the amount of relevant communication between particular students. When clicking on the learner’s node, a pop-up menu displays the topics of the learner’s exchanges as well as the proportion engagement in each topic.

The second type of social learning network in Fig. 11 opens when the learner clicks on a particular topic in the pop-up menu in general discussion network. A new undirected graph represents learners and their exchanges about that topic. Using this analytic, learners can consider whose posts to read or respond to. For example, Nelson struggles to understand linear regression and engages in self-directed study. As every topic has its creator, nStudy considers Jane’s message as a message to the topic’s creator and the edge between these two participants in the graph is weighted as 1 [16]. When Jane submits her next relevant post to the creator or the creator responds directly to Jane, the weight of the edge connecting their nodes increases by 1. In addition, this concept limits the role of a topic’s creator – she can not respond directly to topic s/he created, but can reply to learners’ posts.

Figure 10: Social learning network of hub discussions

![Figure 10: Social learning network of hub discussions](image)

Two types of social learning networks are generated by nStudy. The network in Fig. 10 represents a general discussion across all topics in the hub. Using this graph, the learner and the course instructor can easily identify engaged students and gain better insight to network structure, such as the amount of relevant communication between particular students. When clicking on the learner’s node, a pop-up menu displays the topics of the learner’s exchanges as well as the proportion engagement in each topic.

The second type of social learning network in Fig. 11 opens when the learner clicks on a particular topic in the pop-up menu in general discussion network. A new undirected graph represents learners and their exchanges about that topic. Using this analytic, learners can consider whose posts to read or whom to send a message to gain needed information. For example, Nelson struggles to understand linear regression formulas and therefore cannot create a prediction model. Viewing the “Linear Regression” graph, he observes Jane is not among the most influential participants under this topic. According to the weights of her edges, she engages in modest communication with some of the most influential students under this topic. A list view cannot clearly represent this.

Nelson recalls his successful collaboration with Jane on another project that involved understanding formulas and estimates Jane could explain formulas about linear regression, too. Nelson metacognitively analysed his knowledge gap and forecast Jane’s ability to teach him specific material, so he decides to email Jane asking her for help, even though she is not among the most engaged learners in this topic network.

Discussion Topic: **Linear Regression**

![Figure 11: Social learning network for a specific topic](image)

3. CONCLUSION

To help learners self-monitor their learning processes and be more aware of changes they can make to improve learning, we propose learning analytics that use the ICAP framework. Using trace data and text mining applied to the texts learners are assigned to read and texts they generate, nStudy can classify learners’ engagements in terms of ICAP’s categories and use these classifications to provide meaningful feedback to learners about learning processes. Building on the first step of analysis, learning analytics can support metacognitive engagement in common learning activities including studying, drafting essays and exchanging information with peers. An important aspect of effective analytics that remains to be examined is how to frame learning analytics reports about cognitive engagement so that the learner is motivated to engage in metacognitive monitoring and control, i.e., to self-regulate learning.

REFERENCES


