Toward Integrating Cognitive Tutor Interaction Data with Human Tutoring Text Dialogue Data in LearnSphere

Stephen E. Fancsali
Steven Ritter
Susan R. Berman
Carnegie Learning, Inc.
437 Grant Street, Suite 1906
Pittsburgh, PA 15219, USA
sfancsali@carnegielearning.com
sritter@carnegielearning.com
sberman@carnegielearning.com

Michael Yudelson
Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213, USA
myudelson@cs.cmu.edu

Vasile Rus
Donald M. Morrison
Institute for Intelligent Systems &
Department of Computer Science
University of Memphis
Memphis, TN, USA
vrus@memphis.edu
chipmorrison@gmail.com

ABSTRACT
We present details of a large, novel dataset that includes both Carnegie Learning Cognitive Tutor interaction data and various data related to learner interactions with human tutors via an online chat while using the Cognitive Tutor. We discuss integrating these two data modalities within LearnSphere and propose workflows (and corresponding analyses) germane to our U.S. Department of Defense Advanced Distributed Learning Initiative-funded Integrating Human and Automated Tutoring Systems (IHATS) project, demonstrating various aspects of the potential of the LearnSphere framework.

Keywords
Intelligent tutoring systems, Cognitive Tutor, DataShop, LearnSphere, multi-modal data, human tutoring, mathematics education

1. INTRODUCTION
The Integrating Human and Automated Tutoring Systems (IHATS) project is a U.S. Department of Defense Advanced Distributed Learning Initiative-funded venture led by Carnegie Learning, Inc., in partnership with researchers from the University of Memphis and Carnegie Mellon University. IHATS leverages a unique dataset collected over the period of June to December 2014 from adult learners in two higher education (developmental) algebra courses that featured Carnegie Learning's Cognitive Tutor (CT) [11] intelligent tutoring system (ITS) for mathematics and provided students with the ability to seek human tutoring via an online chat mechanism.

The goal of the IHATS project is to provide insights into what cognitive factors (e.g., error rates) and non/meta-cognitive factors (e.g., behavior like "gaming the system" [1-3] and affective states like boredom [5]) are likely to drive students to seek out human tutoring via the available chat mechanism and what predicts if that tutoring is likely to be successful in driving improved learning outcomes. Insight into predictors and possible determinants of human tutoring use as well as effectiveness of such use will provide a foundation for studying "hand-offs" between automated and human tutoring and instructional modalities. Insights into instructional and tutoring hand-offs may have a variety of practical implications. For example, such insights can inform classroom/instructional best practices (e.g., teachers having an ability to guide students as to when it is best for them to ask their fellow students or the teacher for help versus working with the help affordances of an ITS). Such insights may also guide technical design and feature additions for ITSs like the Cognitive Tutor (e.g., an online recommendation system based on cognitive and non-cognitive factors that drives a student to the right kind of help, whether human or automated).

The multi-modal data we consider in IHATS provide traditional CT ITS learner interaction data of the sort analyzed by a wide variety of educational data mining studies and stored in the Pittsburgh Science of Learning Center's LearnLab DataShop format [9]. In addition to process data about student interactions with the CT, we have both chat transcripts for student interactions with human tutors as well as labels that describe various aspects of these tutoring sessions, which we will now explain.

Our approach to processing tutoring chat transcript data for analysis has been to use human annotators to provide labels, over a sample of 500 transcripts of human tutoring chat sessions, for specific dialogue acts [13] and dialogue modes that occur within these sessions as well as overall assessments of the "educational soundness" and learning effectiveness of these sessions. Machine learning classifiers are trained on these human-annotated labels, and these models are then applied to the large sample of approximately 19,000 transcripts to provide meta-data about each human tutoring chat session.

LearnSphere provides a novel and innovative platform for storage and analysis of these multi-modal educational data. The present paper describes our initial approach to representing these multi-modal data in LearnSphere, details several of the analyses we intend to pursue, and proposes LearnSphere workflows that will enable such analyses as well as future analyses.

2. WORKFLOW METHOD
2.1 Data Inputs
Cognitive Tutor usage data are comprised of approximately 88 million actions (i.e., DataShop transactions) from nearly 5,000 learners using CT in target courses. These data are represented in the PSLC DataShop MySQL format. Additionally, we have (and will have) various annotations (i.e., labels) for each of the human tutoring chat sessions in which learners participated that will be provided as custom fields in the *cf_tx_level_big* table in the DataShop MySQL format. Annotations and tags for each tutoring chat session will be associated with entries (i.e., the

transaction_id) in the *cf_tx_level_big* for the CT transaction that immediately precedes the beginning of the chat session.[1]

Various columns of the *tutor_transaction* table and several related tables would be used in the workflows we propose, including, but not limited to, *outcome*, *duration*, *subgoal_id*, and *problem_id*, as well as columns from tables including *session*, *subgoal*, and *skill* that provide information about the particular knowledge components (KCs) or skills to which student actions/transactions are mapped.

2.2 Workflow Model

LearnSphere provides an intuitive user interface and analytics affordances that may assist in a variety of analyses in the IHATS project. The general goal of our workflow(s) is to produce models of meta-/non-cognitive factors for learners in our dataset, including machine-learned models (i.e., "detectors") of gaming the system behavior [1-3], off-task behavior [4], and affective states like boredom, frustration, and engaged concentration [5]. Such detector models have proven fruitful in illuminating various associations and possible causal relationships among meta-/non-cognitive factors as well as between such factors and learning, for example, between gaming the system behavior and final course exam scores, in a similar population of adult learners using CT in similar algebra courses [8]. These results make us hopeful that such meta-/non-cognitive factors, in addition to cognitive factors like hint use, might help predict whether students turn to human tutoring from the CT and whether/if human tutoring tends to be especially successful under certain conditions.

One possible, omnibus workflow would begin with raw DataShop format data and output the final results of aforementioned detector models (built for a specific product like CT Algebra), including transaction-level predictions of whether particular actions are likely to be instances of gaming the system or off-task behavior or whether learners are likely to be experiencing specific affective states in particular "clips" (i.e., time intervals) of CT usage. However, such a workflow can (should) be broken down into its (more general) constituent components so that elements of the workflow can be generalized to other products, environments, and settings, including other ITSs and educational technologies. For example, Bayesian Knowledge Tracing (BKT) [7] parameter estimates for knowledge components in our dataset are required inputs to later steps in "building" detector models, and LearnSphere already supports a workflow to learn BKT parameter estimates from data.

Detector models for CT also rely on features engineered from fine-grained transaction level data. One component of the LearnSphere suite of tools should include the ability to easily generate such features and enable researchers to easily specify new features to be engineered/specified from fine-grained data. Ideally, features could be engineering to range over a variety of levels of aggregation and units of analysis, including at the transaction, problem, session, and student level (e.g., enabling calculation of average transaction, problem, or session time as well as total student time) and over different time spans (e.g., hint requests before a learner's first interaction with a human tutor or on the problem the learner was working as she initialized a human tutoring chat session). Features used by existing detectors capture facets of fine-grained, transaction-level data that include (but are not limited to): speed with which actions are taken after making at least one error on a problem-solving step in the CT (to detect gaming the system [1-3]), the maximum number of incorrect actions or hint requests for any particular skill within a "clip"/period of problem-solving time (to detect boredom [5]), and many others. Engineered features can then be provided as input to a variety of statistical and machine learning models, including those that comprise existing detector models, but simpler options like linear regression may also be useful.

To enable multi-modal analysis, feature engineering from text data like our chat session transcripts would likely be helpful for a variety of possible analyses. The goal of the IHATS project will be to combine results of machine learned models applied to text data (i.e., to classify dialogue acts, sub-acts, and modes) to determine characteristics of the human tutoring interaction that are associated with improved performance in the CT. Other possibly important, text-based features could also be extracted (e.g., content-related features such as how many content words or domain/topic specific words the student or tutor generated as in [12]). Affordances within LearnSphere workflows could be designed to allow for such analyses using features engineered from both CT usage data and text modalities.

2.3 Workflow Outputs

A variety of possible outputs can result from our proposed workflow(s). BKT parameters for KCs estimated from our large dataset of approximately 88 million learner actions may be of use to EDM and learning analytics researchers, and those models can be evaluated by a variety of metrics available within DataShop/LearnSphere. Transformed datasets with columns corresponding to features engineered within the workflow (and rows corresponding to appropriate units of analysis/aggregation) can be exported for use in statistical and other software tools. Specific tools for the EDM community, including detector models of behavior and affect (based on engineered features and models parameterized/estimated for particular products and systems), could also be included as affordances to workflows. Assuming general analytical tools are incorporated into LearnSphere, statistical models about relationships among such learner behavior and affect (i.e., the output of detectors) could also be the output of a workflow (i.e., using the results of particular models as input to other models in a "discovery with models" approach [6, 8]). For the IHATS project, we are likely to pursue modeling within the framework of algorithmic search for graphical causal models [14] to uncover possible causal relationships among particular behaviors, affective states, learner interactions with human tutoring, and learning. This approach has been fruitfully applied in analyses of CT data from a similar population [8] as well as in other EDM and learning analytics studies (e.g., [10], among others). Such search techniques could, in principle, be made available within LearnSphere workflows (among a bevy of other statistical and machine learning techniques) to provide a "one-stop-(Data)Shop" for applying such analyses to learner data.

3. DISCUSSION

LearnSphere provides an exciting opportunity to make a variety of general interest workflows available to the broader EDM community of researchers. As we have outlined above,

[1] While contractual restrictions forbid us from releasing raw tutoring chat session transcripts, the distributed nature of LearnSphere will allow us to locally (and thus privately) integrate even the chat transcripts, rather than just annotations/tags associated with these transcripts, within a single DataShop style database.

investigators' choice(s) of workflows (and components thereof) are likely to depend on their scientific interests and goals. Rich learner data from environments like ITSs, educational games, and MOOCs can be readily used to make progress on a variety of questions about cognitive modeling and data-driven improvements to student/KC models as well as to questions about relationships between cognitive, non-cognitive, and meta-cognitive factors at play as learners interact with such systems. LearnSphere pushes the boundaries of the types of multi-modal data with which researchers will be able to more easily work, including human tutoring chat transcripts (and corresponding meta-data) in the IHATS project, to pursue innovative research in the learning sciences. LearnSphere's making analyses of such learner data more readily generalizable and replicable will be a great service to the learning sciences and EDM community.

4. ACKNOWLEDGMENTS

The IHATS project is funded by the U.S. Department of Defense, Advanced Distributed Learning Initiative, Contract W911QY-15-C-0070.

5. REFERENCES

[1] Baker, R.S.J.d., Corbett, A.T., Koedinger, K.R., Evenson, E., Roll, I., Wagner, A.Z., Naim, M., Raspat, J., Baker, D.J., Beck, J. 2006. Adapting to when students game an intelligent tutoring system. In *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (Jhongli, Taiwan, 2006). 392-401.

[2] Baker, R.S., Corbett, A.T., Koedinger, K.R., Wagner, A.Z. 2004. Off-task behavior in the Cognitive Tutor classroom: when students "game the system." In *Proceedings of ACM CHI 2004: Computer-Human Interaction* (Vienna, Austria, 2004). 383-390.

[3] Baker, R.S.J.d., Corbett, A.T., Roll, I., Koedinger, K.R. 2008. Developing a generalizable detector of when students game the system. *User Model. User-Adap.* 18 (2008), 287-314.

[4] Baker, R.S.J.d. 2007. Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of ACM CHI 2007: Computer-Human Interaction* (San Jose, CA, April 28 – May 3, 2007). ACM, New York, 1059-1068.

[5] Baker, R.S.J.d., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G.W., Ocumpaugh, J., Rossi, L. 2012. Towards sensor-free affect detection in Cognitive Tutor Algebra. In *Proceedings of the 5th International Conference on Educational Data Mining* (Chania, Greece, June 19-21, 2012). International Educational Data Mining Society, 126-133.

[6] Baker, R.S.J.d., Yacef, K. 2009. The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining* 1 (2009), 3-17.

[7] Corbett, A.T., Anderson, J.R. 1995. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Model. User-Adap.* 4 (1995), 253-278.

[8] Fancsali, S.E. 2014. Causal Discovery with Models: Behavior, Affect, and Learning in Cognitive Tutor Algebra. In *Proceedings of the 7th International Conference on Educational Data Mining* (London, UK, July 4-7, 2014) International Educational Data Mining Society, 28-35.

[9] Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., Stamper, J. 2011. A data repository for the EDM community: the PSLC DataShop. In *Handbook of Educational Data Mining*, C. Romero, S. Ventura, M. Pechenizkiy, & R.S.J.d. Baker, Eds. CRC, Boca Raton, FL.

[10] Koedinger, K.R., Kim, J., Jia, J.Z., McLaughlin, E.A., Bier, N.L. 2015. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the Second (2015) ACM Conference on Learning@Scale* (Vancouver, Canada, March 14-15, 2015). ACM, New York, 111-120.

[11] Ritter, S., Anderson, J.R., Koedinger, K.R., Corbett, A.T. 2007. Cognitive Tutor: applied research in mathematics education. *Psychon. B. Rev.* 14 (2007), 249-255.

[12] Rus, V., Stefanescu, D. 2016. Non-intrusive assessment of learners' prior knowledge in dialogue-based intelligent tutoring systems. *International Journal of Smart Learning Environments* 3 (2016), 1-18.

[13] Searle, J.R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge UP.

[14] Spirtes, P., Glymour, C., Scheines, R. 2000. Causation, Prediction, and Search. 2nd Edition. MIT, Cambridge, MA.