Using Temporal Association Rule Mining to Predict Dyadic Rapport in Peer Tutoring

Michael Madaio
Carnegie Mellon University
Pittsburgh, Pennsylvania
mmadaio@cs.cmu.edu

Rae Lasko
Carnegie Mellon University
Pittsburgh, Pennsylvania
rlasko@andrew.cmu.edu

Amy Ogan
Carnegie Mellon University
Pittsburgh, Pennsylvania
aeo@cs.cmu.edu

Justine Cassell
Carnegie Mellon University
Pittsburgh, Pennsylvania
justine@cs.cmu.edu

ABSTRACT

Social relationships, such as interpersonal closeness or rapport, can lead to improved student learning, but such dynamic, interpersonal phenomena can be difficult for educational support technologies to detect. In this paper, we describe an approach for rapport detection in peer tutoring, using temporal association rules learned from nonverbal, social, and on-task verbal behaviors. From a corpus of 60 hours of annotated multimodal peer tutoring data, we learn the temporal association between behaviors and the rapport score for each 30-second "thin-slice". We then train a stacked ensemble classification model on those association rules and evaluate our ability to reliably predict rapport using multimodal behavioral data. We find that our approach allows us to predict rapport well above chance, and more accurately than two baseline models. We are able to predict high rapport more accurately for strangers and low rapport more accurately for friends, which we believe holds promise for the integration of rapport detection into collaborative learning supports and intelligent tutoring systems.

Keywords

rapport, association rule mining, peer tutoring, social states

1. INTRODUCTION

Social relationships, such as the long-term closeness of friends or the short-term rapport built while getting to know someone, have been shown to result in benefits for student learning, such as increased help-seeking, productive cognitive conflict, and elaborated reasoning [2]. In collaborative learning settings, higher interpersonal rapport between students is associated with productive educational processes such as instances of transactive reasoning [13] and greater learning gains over time [18]. Educational technologies, such as intelligent tutoring systems (ITS) and pedagogical agents, increasingly attempt to reap the benefits of interpersonal closeness and rapport between humans and agents to improve engagement, motivation, or trust in the pedagogical agent [19]. However, before educational technologies can respond appropriately to the rapport between collaborating students, or build rapport between students and a pedagogical agent, they must first model that rapport as it changes over time, given the available behavioral data. The educational data mining community has developed, over the last several decades, detectors of individual student phenomena such as frustration, boredom, engagement, carelessness, and many others [3, 17], but it has developed relatively fewer methods for modeling inter-personal social phenomena such as the rapport between members of a collaborative group or peer tutoring dyad.

This paper is intended to contribute to the detection of inter-personal social states, such as rapport, through nonverbal, task (verbal) and social (verbal) channels, captured through audio and video input. In this paper, we describe a process for using temporal association rule mining to learn patterns of behaviors from an annotated corpus of nearly 60 hours of dyadic peer tutoring interactions. We then use those temporal association rules to predict the "thin-slice" dyadic rapport level for every 30-second time-slice, via a stacked ensemble model. We find that temporal rules generated from annotations of students' nonverbal, on-task, and off-task social behaviors were overall able to predict rapport at levels well above chance, and at nearly double the prediction performance (AUC) of a baseline approach. We found that this approach allows us to predict high rapport significantly better than low rapport overall, while predicting high rapport for strangers more accurately than for friends.

This paper contributes to the Educational Data Mining (EDM) community in several ways: (1) We describe a process for automatically learning temporal association rules from annotations of nonverbal, and social and on-task verbal behaviors, and using those rules to predict rapport in a stacked ensemble model, compared to two baseline approaches. (2) We describe the variation in the number of high-confidence rules learned for each of the behavioral channels, to inform future developers of rapport detectors of the

data sources that may be most fruitful to capture. (3) We evaluate the predictive performance of those temporal rules in predicting rapport for both friends and strangers, thereby addressing both short- and long-term rapport.

2. RELATED WORK

In order to choose the behaviors used to predict rapport, we draw on a framework of rapport-building proposed by [22]. In this theoretical model, rapport is a dyadic phenomenon, co-constructed over time by both members of the dyad. According to [22], rapport is developed through nonverbal behaviors and verbal social conversational strategies that serve various social functions and sub-goals in rapport development, such as face management, mutual attentiveness, and coordination [22]. Our work extends [22]'s approach by also incorporating the task-related verbal strategies from both tutor and tutee, such as feedback, instructions, and task-related questions which are essential for the tutoring process, and which we hypothesize will impact, and be impacted by the rapport between members of a peer tutoring dyad [9].

Prior researchers in discourse analysis, multi-modal interaction, and dialogue systems have developed detectors for various aspects of interpersonal relationship development, such as Yu et al.'s friendship prediction for peer tutoring dyads, which found that dyadic features such as mutual gaze and smile behaviors were predictive of friendship [21]. In prior EDM work, some [15] have used the temporal co-occurrence of nonverbal behaviors (operationalized as Facial Action Units) to capture "behavioral synchronicity" in collaborative problem-solving dyads. Others have developed automatic classifiers of on-task-related interpersonal behaviors, such as [14]'s method for classifying socio-cognitive conflict in collaborative learning within an intelligent tutoring system. Others, such as [20], have developed automatic classifiers of dyadic impoliteness and positivity, work that we build on here with the social conversational strategies we incorporate into the association rules. Prior work has demonstrated the effectiveness of out-of-domain social talk in pedagogical agents, such as [8]'s social pedagogical agent used in collaborative learning.

2.1 Temporal Patterns in Behavior

As rapport-building is a dynamic phenomena, it is impacted by the contingent patterns of verbal and nonverbal behavior. Ohlssen et al. describe how popular methods for discourse analysis that use a "code-and-count" method [12] collapse the temporal dimension and are thus unable to understand the rich patterns of interaction likely to impact learning, or rapport. To address this gap, we draw on the Temporal Interval Tree Association Rule Learning (Titarl) framework [7] to discover temporal patterns of verbal and nonverbal behavior and their association with the dyadic rapport between members of a tutoring dyad for every 30-second time slice. The Titarl framework has been previously used to analyze medical patients' vital sign data [7], and in our lab, [24] have used Titarl to identify patterns of social conversational strategies and nonverbal behaviors predictive of levels of rapport. Crucially, however, [24] did not include the tutoring and learning behaviors that are the heart of the task component of the peer tutoring interactions, and which are likely to impact rapport through, for example, the face-threatening nature of providing feedback or instruc-

tions [9]. Therefore, in order to more effectively predict the rapport between members of a peer tutoring dyad, we include rules learned from the nonverbal, social verbal, and tutoring-related verbal behaviors.

3. METHODS

3.1 Research Questions

RQ1: Can temporal association rules learned from social conversational strategies, task, and nonverbal behaviors in peer tutoring be used to predict rapport at levels above chance? From [7] and [24], we believe that they can, and that we can improve the predictive performance by adding the task-related verbal behaviors.

RQ2: Is a classifier trained on temporal association rules better able to predict rapport (a) for some relationship types than others or (b) at some levels than others? Following [24], we believe we will be better able to predict high rapport among strangers than among friends.

RQ3: Are temporal association rules (TAR) generated using all three channels of on-task (verbal), social (verbal), and nonverbal behavioral better able to predict rapport than rules generated from any one or two of those behavior types? From [9, 21, 24], we believe that including task, social, and nonverbal together will perform the best.

3.2 Data Collection and Dialogue Corpus

The dialogue corpus described here was collected as part of a larger study on the effects of rapport-building on reciprocal peer tutoring [9, 10, 18, 22]. The participants were assigned to 12 dyads that alternated tutoring each other in Algebra for 5 weekly hour-long sessions, for a total of 60 hours. Half were male and half were female, assigned to same-gender dyads. To investigate how the impact of various task, social, and nonverbal behaviors on rapport differs between dyads with varying degrees of interpersonal closeness, we used friendship as a proxy for long-term rapport and thus asked half of the participants to bring a same-age, same-gender friend to the session with them, and for the other half of the dyads, we paired them with a stranger, using the 5 weeks to capture short-term rapport-building. Audio and video data were recorded, transcribed, and segmented for clause-level dialogue annotation.

3.3 Thin-Slice Rapport Ratings

The rapport between the participants, was evaluated using a 'thin-slice' approach [1]. First, the corpus was divided into 30-second video slices, then shuffled (so the raters did not inadvertently rate the change in rapport), and provided to naive, third-party raters. Three such raters rated the rapport present in each slice on a Likert scale from 1-7, from lowest possible rapport to highest possible rapport. A single rating was then chosen for each slice using an inverse bias-corrected weighted majority vote approach, described in [18], to account for potential over-use or under-use of certain labels by the raters. The final consensus measure of inter-rater reliability, or Cronbach's α, was .86, justifying the use of this rating selection method [18]. This rating was used when learning the associations between the task, social, and nonverbal behaviors and the rapport level.

Table 1: Annotation Types, Labels, Definitions, and Examples

Type	Label	Definition	Example
Task	Knowledge-telling	Stating procedures or the answer	Divide it by 9.
Task	Knowledge-building	Providing explanations	That's because it can be reduced.
Task	Correct or Incorrect Feedback	Evaluating their partner's correctness	No, that's not quite it.
Task	Shallow Question	Asking about procedures or answers	Is that right?
Task	Deep Question	Asking about reasoning or concepts	Why would you do that?
Social	Self-Disclosure	Sharing personal information about oneself	I suck at negative numbers.
Social	Refer to Shared Experience	Discussing an experience they had together	Remember that soccer game?
Social	Violation of Social Norms	Statements that break social conventions	It's a zero, dummy.
Social	Praise	Positive acknowledgment of the other	You're so smart!
Social	Reciprocation	Responding to a conversational move with the same conversational move.	Tutor self-discloses, then the tutee self-discloses

3.4 Dialogue Annotation

To investigate the impact of rapport-building verbal (social and task) and nonverbal behaviors, we annotated our dataset for 3 types of nonverbal behaviors, 5 types of social conversational strategies, and 5 types of tutoring and learning behaviors, as shown in Table 1, all annotated with $> .7$ Krippendorff's α. The nonverbal behaviors annotated were head nods, smiles, and shifts in eye gaze from the partner, to the Algebra worksheets, to anywhere else, similar to [21]. The social verbal behaviors were chosen according to [22]'s theory of rapport-building, behaviors such as self-disclosure, reference to shared experiences, violation of social norms, and others. The on-task verbal behaviors annotated are based in part on [16]'s work on knowledge-telling and knowledge-building, as well as [6] work with procedural and conceptual questions, described in more detail in [10].

3.5 Temporal Association Rule Mining

To investigate the impact that these nonverbal, task, and social behaviors had on rapport at a 30-second thin-slice level, we adopted a temporal association rule mining approach, following [23]. The framework we use, the "Temporal Interval Tree Association Rule Learning" (Titarl) algorithm [7], allows us to identify temporal patterns of behaviors within each time slice that are probabilistically associated with the value of rapport for that slice. For each 30-second time window, a rule is learned much like the generic rule below.

"If event A happens at time t, there is 50% chance of event B happening between time t+3 to t+5".

Our data is comprised of both multivariate symbolic time sequences (the nonverbal, task, and social behaviors) and multivariate scalar time series (the rapport value for each slice). The Titarl algorithm will learn a large set of rules on a subset of our data (the training set), filter those rules based on a set of parameter thresholds, fuse similar simple rules into more complex rules, which we then use in predicting rapport on a held-out test set. Because we believed that the ways that friends and strangers build rapport with each other over 5 weeks are likely to differ following [23], we ran the Titarl algorithm on sets of friend dyads and sets of stranger dyads separately.
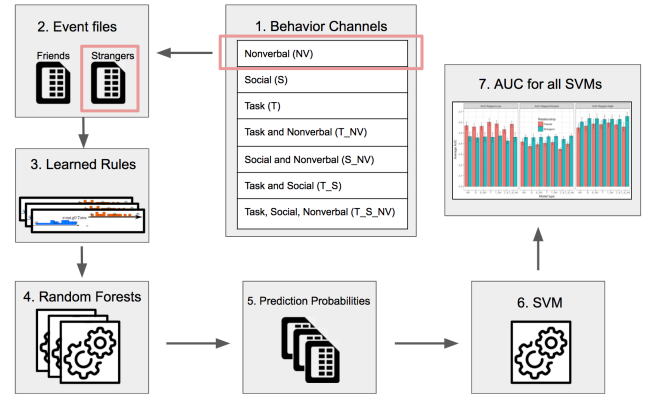
3.6 Rapport Detection Process

Figure 1: Multi-step process for prediction of rapport using temporal association rule mining and a stacked classifier ensemble.

We describe here an approach laid out in Figure 1. We first divided our 6 friend dyads and 6 stranger dyads, with 5 sessions per dyad, into a training set of 4 dyads (20 total sessions) and a held-out test set of 2 dyads (10 total sessions) for both relationship types. Then, (Step 1) we created seven combinations of the Social, Task, and Nonverbal annotations described in Table 1, to identify differences in prediction performance for the different behavior types (RQ3). Next (Step 2), for each of those behavior combinations, we created a matrix M with $n+1$ columns, with $n =$ the total number of annotation types (used by the tutor and the tutee), described in Table 1, with the first column in M being the start time, in seconds, of each behavior. Each row in M was an event, or the start of an annotated verbal or nonverbal behavior. From each matrix M, we generated an "event file" which included the behavior sequence as well as the scalar time series of the rapport value for the 30-second time slice within which those behaviors occurred.

Then, using these files, we (Step 3) learned a set of association rules R for each training set, using the Titarl algorithm [7]. These rules contain a head, which is the scalar output value of rapport (an integer from 1-7), and a body, which is the ordered set of annotated behaviors used to predict the rapport in each slice. Prior to learning, we specified the minimum confidence (the probability of the prediction of the rule to be true) at 50%, the minimum support (the per-
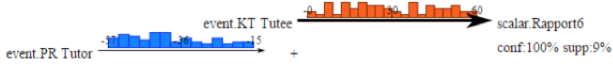
Figure 2: Example temporal association rule for strangers with high rapport, with 100% confidence, 9% support, and 44 uses.
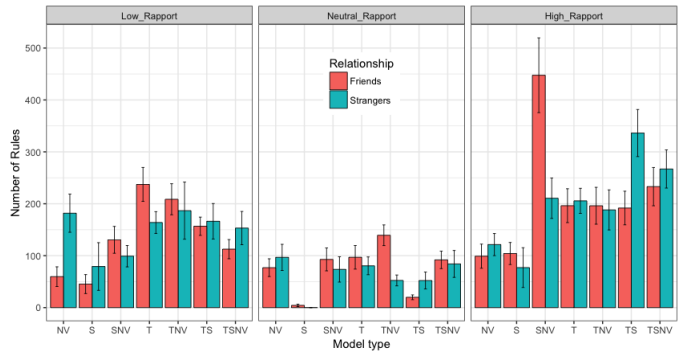

Figure 3: Mean number (and standard error) of rules learned from 7 behavioral channels, with high confidence, support, and usage, for friends and strangers with low, neutral, and high rapport.

centage of events explained by the rule) at 5%, and the minimum number of uses for each rule at 10, following [24]. An example of a rule can be seen in Figure 2, where a Tutor's use of Praise (PR), followed by the Tutee's "Knowledge-telling" (KT), or self-explanation, is associated with a rapport value of 6 (high), with confidence of 100%, support of 9%, and 44 uses in that model. This rule was learned from a Task and Social behavioral model, for a dyad of Strangers. The nature of these data can be further illustrated with another example, from a highly confident association rule learned from the Task and Social model, for dyads of Friends. The following high-confidence rule is associated with Rapport of 1 (low): a Tutee asks a Shallow Question, receiving four "Knowledge-telling" utterances in a row from their Tutor, to which the Tutee responds with a "Social Norm Violation". In other words, the tutee asks about the procedure, the tutor tells him what to do in multiple utterances, and the tutee responds with some norm violation, perhaps rudeness. To ensure that the rules learned from each set of dyads were not overfit to the particular training set of dyads used to learn them, we learned a rule set (i.e. repeated Steps 1-3) for all possible combinations of the 6 friend and 6 stranger dyads, resulting in 15 "folds" for friend dyads and 15 for strangers (i.e. choosing all possible sets of 4 dyads to use as training sets from the 6 total dyads). Each fold had several hundred association rules learned above our threshold for confidence, support, and usage. In Figure 3, we show the mean number of rules learned, showing only those with confidence, support, and usage above the median for ease of visualization.

After learning the rules, in Step 4 we use the rules to train random forest classifiers to predict the rapport level for each 30-second slice. To do this, we first generated a matrix N for each rule set in each of the 30 training sets, with a row for each rule event in that set, and $n+1$ columns, where n is the number of rules in that train set, and the final column was a binary indicator of the rapport value for that time slice. We ran 7 random forest classifiers (one for each rapport level) for each matrix N, for each of the 15 folds of friends and 15 folds of stranger training sets, giving us (in Step 5) a prediction probability estimate for each of the 7 rapport values, for each event in every fold, for each of the 7 behavioral channels (from Step 1). Finally, we wanted to evaluate the relative impact of those 7 behavior types, and so we composed different combinations of nonverbal, social, and task behavior. We then, in Step 6, use the prediction probability output by the random forest classifiers as the input features in training a single multi-class Support Vector Machine (SVM) classification model for each of the 30 folds to predict the overall rapport level for each time slice in that fold. In the following section, we discuss the performance of this final classification step in predicting rapport for each relationship type and evaluate its performance against two baselines from earlier steps in the process.

4. RESULTS

First, before investigating our first research question about the performance of our approach in predicting rapport, we wanted to inspect the total number of rules learned from each behavioral channel with high confidence, support, and usage, to better understand the extent to which the number of highly confident temporal rules varied for each behavioral type. See Figure 3 for the mean number and standard errors for rules learned above the median confidence, support, and usage for low, neutral, and high rapport for friends and strangers. Based on the distribution of slices at each level, we converted the 7 scalar rapport values to the low (1-3), neutral (4), and high (5-7) rapport levels.

We see that friends had significantly more (t(20.8)=2.7, p=.01) high-confidence Social and Nonverbal ("SNV") rules learned in High Rapport slices than the next largest behavioral channel, the "TSNV" channel, combining Task, Social, and Nonverbal. This suggests that a method for detecting high rapport between friends that uses Social and Nonverbal behaviors will have many more high-confidence, frequently occurring rules with which to predict rapport than using other sets of behavior types. Conversely, for rules learned from Friend dyads for Low Rapport slices, there is a significantly (t(26.6)=2.6, p<.05) greater number of high-confidence, frequently occurring Task ("T") rules than rules learned from the Social and Nonverbal (SNV) behaviors. That is, there are substantially more high-confidence, high-support, and frequently occurring ways in which Friends displayed Low Rapport through their on-task behavior (and on-task combined with nonverbal, "TNV") than through other available channels. This suggests that a method for detecting students' low rapport, for a dyad of friends, may benefit from incorporating the task-related behaviors such as instructions, explanations, questions, and provision of feedback in addition to purely social behaviors, as in [23]. Similarly, for Strangers, their Task and Social ("TS") channel had the largest number of rules learned associated with High Rapport slices, significantly more than the "SNV" behaviors (t(26.8)=1.2, p<.05), though not significantly more than the TSNV behaviors. This suggests that a detector of high rapport that leverages Task and Social behaviors may have more high-confidence association rules from which to draw for its

Table 2: Average PR-AUC (and standard deviation) of 3 rapport prediction models

Model	PR-AUC
IF Baseline	.42 (.07)
RF Baseline	.33 (.03)
TAR Ensemble	.60 (.08)

classification of high rapport for students without a prior friendship relationship (i.e. "strangers") than one relying solely on Strangers' social and nonverbal behaviors.

Then, to evaluate the overall performance of our approach in predicting low, neutral, and high rapport, we used the prediction probability from the 7 binary random forest classifiers (from Step 5) as the input into a 3-way one-vs-rest SVM classifier, for every behavioral channel model (Step 6). We first ran a 10-fold cross-validated grid search on our training set to discover the optimal set of parameters to use for the SVM model, using an RBF kernel, with $C=10$ and $\gamma = 1$. From the SVM, we use the average area under the Precision-Recall curve (PR-AUC) for each of the 7 behavioral models as our performance measure, following [5].

First, for RQ1, to validate the appropriateness of our stacked ensemble approach ("TAR Ensemble"), we compare its prediction performance to two baseline approaches. We compare first to a baseline that treats the annotated behaviors in each slice as independent features in an SVM using the same parameters ("IF Baseline"). The TAR Ensemble significantly ($t(413) = 24.4$, p<.001) outperforms the IF Baseline with a mean AUC of .60 (sd = 08) for the TAR Ensembles, compared to a mean AUC of .42 (sd = .07) for the IF Baseline. We then compare the TAR Ensemble to another baseline ("RF Baseline") that simply takes the largest prediction probability from the 7 random forests (Step 5 in Figure 2) as the predicted class value, using random selection for ties. The TAR Ensemble significantly ($t(256) = 46$, p<.001) outperforms the RF Baseline by nearly 2 to 1, with a mean AUC of .60 (sd = 08) for our approach and a mean AUC of .33 (sd = .03) for the RF Baseline. See Table 2 for a summary of the PR-AUC values for each model.

For RQ2a, we find that the Stacked Ensemble is better able to predict High Rapport than Low ($t(417)=5.9$, p<.005). For RQ2b, we are better able to predict Low Rapport for Friends than Strangers ($t(197) = 5.8$, p<.001). Conversely, we are better able to predict the rapport among Strangers than among Friends for both Neutral ($t(206.5) = 5.5$, p<.001) and High rapport levels ($t(207) = 2.7$, p<.01). For RQ3, no single set of behavioral channels significantly outperformed the others, in an ANOVA of the PR-AUC measure with each relationship type (Friend/Stranger), rapport level (Low/Neutral/High), and behavioral type (TS, TSNV, etc).

5. DISCUSSION AND CONCLUSION

Interpersonal social dynamics provide the grounding for learning interactions, whether students are learning collaboratively, in peer tutoring, or working with their classroom teacher or even a virtual agent. However, technological supports for learning often focus on detecting and modeling individual, intra-personal states such as students' affect or

engagement, without considering the latent social state underpinning their interactions with others. In this work, we present one method for detecting the latent social state of interpersonal rapport in learning interactions, using a temporal association rule mining approach to learn patterns of nonverbal and verbal (social and task) behaviors, as input in predicting the rapport level in a stacked ensemble model. Our ensemble approach outperforms two baselines, (1) the independent behaviors as features, and (2) the random forests trained on the temporal association rules.

We find that, overall, our approach is better able to predict high rapport than low rapport, and it predicts high and neutral rapport more accurately for Strangers than for Friends, while predicting low rapport more accurately for Friends than for Strangers. This is good news for designers of virtual agents that want to detect and build rapport with a new student, or designers of computer-supported collaborative learning technologies that want to detect rapport in learning. However, contrary to our expectations (for RQ3), we saw no significant difference in prediction performance across the models generated from different combinations of behavior types (e.g. SNV, TSNV, etc). We did see a significant difference in the total number of association rules learned from those behavior types, however, suggesting that rapport detectors will be better able to predict rapport if they use the behavior types that occur more frequently in learning. For instance, a rapport detection method for strangers that incorporates Task and Social behavior will have significantly more high-confidence, high-support association rules with which to detect the rapport between them.

One of the limitations of this current approach is that, while it may reach quite good levels of performance in detecting rapport, the large number of rules learned make it difficult to identify the specific rules that are most predictive of rapport, in addition to concerns about dimensionality. This work is limited by the small sample size, and by being restricted to same-gender dyads; using a larger set of dyads to conduct these analyses may reveal differences in prediction performance for different behavioral types (social, task, nonverbal), if they exist. We have currently finished collecting 22 dyads' worth of interactions among strangers (over 40 hours), and we will be conducting a similar set of annotations and analyses on them. In this paper, the thin-slice rapport ratings and annotations were hand-annotated from a corpus of audio/video data, limiting the automaticity of this approach. However, we are in the process of moving to a crowd-sourced method for obtaining the ground truth rapport ratings for each 30-second slice. Preliminary experiments for crowd-sourcing the thin-slice rapport annotation using Amazon Mechanical Turk have yielded a Krippendorff's α of 0.69 across 3 raters for each thin-slice.

In future work, we intend to use this rapport estimation method for a rapport-building virtual agent in an intelligent tutoring system. We have developed automatic classifiers for the three types of nonverbal behaviors described here, using the OpenFace system [4], and social conversational strategy classifiers, such as those described by [23], classifiers which have already been integrated into a "socially aware robot assistant" (SARA), as described by [11]. Our next step is to develop a task-related classifier, perhaps similar to that

used in [14], to recognize students' task-related utterances as part of the rapport estimation and reasoning about natural language response generation. We believe that this paper contributes to the larger goal of educational data mining by demonstrating one approach to using multimodal data to model latent social phenomena important to learning, in this case the interpersonal rapport in peer tutoring.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, 111(2):256–274, 1992.

[2] M. Azmitia and R. Montgomery. Friendship, transactive dialogues, and the development of scientific reasoning. *Social development*, 2(3):202–221, 1993.

[3] 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.

[4] T. Baltrušaitis, P. Robinson, and L.-P. Morency. Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), year=2016, organization=IEEE*.

[5] J. Davis and M. Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd ICML*, pages 233–240. ACM, 2006.

[6] A. C. Graesser, N. K. Person, and J. P. Magliano. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied cognitive psychology*, 9(6):495–522, 1995.

[7] M. Guillame-Bert and A. Dubrawski. Learning temporal rules to forecast events in multivariate time sequences. In *2nd Workshop on Machine Learning for Clinical Data Analysis, Healthcare and Genomics. NIPS*, 2014.

[8] R. Kumar, H. Ai, J. L. Beuth, and C. P. Rosé. Socially capable conversational tutors can be effective in collaborative learning situations. In *International Conference on Intelligent Tutoring Systems*, pages 156–164. Springer, 2010.

[9] M. Madaio, J. Cassell, and A. Ogan. The impact of peer tutors ' use of indirect feedback and instructions. In *Computer-Supported Collaborative Learning Conference*, 2017.

[10] M. A. Madaio, A. Ogan, and J. Cassell. The effect of friendship and tutoring roles on reciprocal peer tutoring strategies. In *International Conference on Intelligent Tutoring Systems*. Springer, 2016.

[11] Y. Matsuyama, A. Bhardwaj, R. Zhao, O. J. Romero, S. A. Akoju, and J. Cassell. Socially-aware animated intelligent personal assistant agent. In *17th Annual Meeting of SIGDIAL*, page 224, 2016.

[12] S. Ohlsson, B. D. Eugenio, B. Chow, D. Fossati, X. Lu, and T. C. Kershaw. Beyond the code-and-count analysis of tutoring dialogues. *AIED: Building technology rich learning contexts that work*, 158:349, 2007.

[13] J. Olsen and S. Finkelstein. Through the (thin-slice) looking glass : An initial look at rapport and co-construction within peer collaboration. In *Computer-Supported Collaborative Learning Conference*, in press.

[14] D. Prata, R. Baker, E. Costa, C. Rose, and Y. Cui. Detecting and understanding the impact of cognitive and interpersonal conflict in computer supported collaborative learning environments. In *Educational Data Mining 2009*, 2009.

[15] V. Ramanarayanan and S. Khan. Novel features for capturing cooccurrence behavior in dyadic collaborative problem solving tasks. In *Educational Data Mining 2009*, 2016.

[16] R. D. Roscoe and M. T. Chi. Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *RER*, 77(4):534–574, 2007.

[17] M. O. Z. San Pedro, R. S. d Baker, and M. M. T. Rodrigo. Carelessness and affect in an intelligent tutoring system for mathematics. *IJAIED*, 24(2):189–210, 2014.

[18] T. Sinha and J. Cassell. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And inflLuence*, pages 13–20. ACM, 2015.

[19] E. Walker and A. Ogan. We're in this together: Intentional design of social relationships with aied systems. *IJAIED*, 26(2):713–729, 2016.

[20] W. Y. Wang, S. Finkelstein, A. Ogan, A. W. Black, and J. Cassell. Love ya, jerkface: using sparse log-linear models to build positive (and impolite) relationships with teens. In *Proceedings of SIGDIAL*, pages 20–29. Association for Computational Linguistics, 2012.

[21] Z. Yu, D. Gerritsen, A. Ogan, A. W. Black, and J. Cassell. Automatic prediction of friendship via multi-model dyadic features. In *Proceedings of SIGDIAL*, pages 51–60, 2013.

[22] R. Zhao, A. Papangelis, and J. Cassell. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *IVA*, pages 514–527. Springer, 2014.

[23] R. Zhao, T. Sinha, A. W. Black, and J. Cassell. Automatic recognition of conversational strategies in the service of a socially-aware dialog system. In *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 381, 2016.

[24] R. Zhao, T. Sinha, A. W. Black, and J. Cassell. Socially-aware virtual agents: Automatically assessing dyadic rapport from temporal patterns of behavior. In *International Conference on Intelligent Virtual Agents*, pages 218–233. Springer, 2016.