A Model-based Evaluation of Trust and Situation Awareness in the Diner's Dilemma Game

John O'Donovan
Department of Computer Science
University of California
Santa Barbara
Email: jod@cs.ucsb.edu

Rashaad E.T. Jones
S.A. Technologies Inc.
Marietta, GA 30066
Email: rashaad.jones@satechnologies.com

Laura R. Marusich
Human Research and Engineering Directorate
U.S Army Research Laboratory
Aberdeen Proving Ground, MD
Email: laura.r.marusich.ctr@us.army.mil

Yun Teng
Department of Computer Science
University of California
Santa Barbara
Email: yun-cs@cs.ucsb.edu

Cleotilde Gonzalez
Department of Social and Decision Sciences
Carnegie Mellon University
Email: coty@cmu.edu

Tobias Höllerer
Department of Computer Science
University of California
Santa Barbara
Email: holl@cs.ucsb.edu

***Abstract*—This paper describes a machine learning approach to evaluate the relationship between trust behavior and Situation Awareness (SA) in the context of a 3-player Iterated Diner's Dilemma game. Our experimental setup consisted of a set of 24 supervised studies in which participants played against computer opponents with different cooperation strategies. Three user interfaces were evaluated in the study, each corresponding to a specific level of SA (Perception, Comprehension, and Projection). Two concepts of trust are explored in the study: *Empirical trust behavior* was recorded as the degree of cooperation imparted by the participant in a given condition, and *Self-reported trust* assessments were also collected at regular intervals throughout the study. To explore the relations that exist between these two concepts of trust, and different SA Level conditions, a machine learning approach was applied to train a variety of models to accurately predict trust behavior in each condition. Our best performing algorithm was a J48 rule-based learner, which leveraged SA Level, strategy, and self-reported survey data to predict trust behavior to 76% accuracy, and 67% accuracy using only SA Level and opponent strategy. This is a relative increase of 43% and 24% respectively over a benchmark majority class predictor. These results indicate that support at a level of SA and opponent strategy are good predictors of trust behavior in the Iterated Diner's Dilemma game. Our results also show that trust-prediction models perform best when they are trained on a combination of self-reported data, and strategy / SA Level information.**

I. Introduction

Recent research has focused on the complex relationship between Trust and Situation Awareness (SA) in a variety of contexts [19], building on trust research covering simple recommendation scenarios [14], abstract games [19] and SA research involving complex representations of real world military and crisis-management scenarios [10], [11]. This paper approaches the problem of understanding trust and SA from a machine learning perspective. We posit that probabilistic models can be trained to discover patterns in a history of behavioral data in a variety of different SA contexts, and that these models can be used to accurately predict future trust behavior given an SA level and related context. To evaluate our hypothesis, we test a range of model-based learning algorithms on a data set collected from a study described in [19].

While it has been shown that trust plays an important role in the social dynamics of a network [14], [8], and that trust can be successfully modeled in a variety of contexts [12], [13], [4] a precise relationship between trust and SA Level has yet to be defined in a generalizable manner. To address this challenge, we describe a study of 24 participants who played a version of the 3-Player Iterative Diner's Dilemma game against automated opponents. Participants were given one of three SA Level visualizations to inform them of the game state and opponent behavior. The study evaluated two aspects of trust: *behavioral trust*, measured by the degree of cooperation observed for each participant in the game, and *self-reported trust*, measured through iterative questionnaires administered throughout the game.

II. Related Work

A. Situation Awareness

Situation Awareness (SA) can be thought of as an internalized mental model of the current state of an operators environment the many streams of incoming data, the external surroundings, and other concerns must be brought together into an integrated whole [10]. This unified picture form s the central organizing feature from which all decision-making and action takes place [10]. Formally, SA is defined as the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future [7]. More simply stated, SA involves being aware of what is happening around you to understand how information, events, and your own actions will affect your goals and object ives, both now and in the near future. Research indicates that SA is a fundamental construct driving human decision-making in complex, dynamic environments [5], [6], [7], [17], [7]. These works motivate the idea that effective decision-making requires developing and maintaining SA at three levels, as described below.

- Level 1 SA (Perception) utilizes the processes of monitoring, cue detection, and simple recognition, leading to an awareness of multiple situational elements (objects, events, people, systems, environmental factors) and their current states (locations, conditions, modes, actions),
- Level 2 SA (Comprehension) involves the processes of pattern recognition, interpretation, and evaluation to integrate Level 1 SA elements to understand how this information will impact goals and objectives, and
- Level 3 SA (Projection) is achieved through integrating Level 1 and 2 SA information and extrapolating this information to project future actions and states of the elements in the operational environment .

B. *Trust and Situation Awareness in the Diner's Dilemma*

Decision-makers in social interactions are often unaware of the way their own actions influence other people, and vice versa. For example, in the intensifying environmental crisis, individuals may not realize that their own conservation or pollution can inspire others to behave similarly by activating norms of reciprocity and punishment, respectively. Researchers often use simplified representations of real world social interactions, presenting explicit actions of two players and the corresponding outcomes obtained when players select that action. Probably the most well-known representation of social interactions is the Iterated Prisoners Dilemma. This is a common and simplified representation of a dilemma between two players, who decide to take actions without communication. They must decide whether to cooperate or defect, with defection leading to higher outcomes for each, regardless of the others action, but mutual cooperation leading to higher joint outcomes than mutual or unilateral defection [15]. The main dilemma then is one of avoiding the temptation of short-term defection and instead cooperating, which leads to good mutual long-term outcomes. The proportion of mutual cooperation in this game is often considered a measure of "trust" between the two individuals, a notion that we also explore here.

The Diner's Dilemma is an n-player Prisoner's Dilemma. The scenario is that several individuals go out to eat with the prior agreement to share the bill equally. Each individual will make the decision of whether to order the expensive dish (e.g., lobster) or inexpensive dish (e.g., hot dog). It is presupposed that the expensive dish is better than the cheaper one, but not worth paying the difference when dining alone. The overall best dining experience (food enjoyment divided by price) is achieved when everyone chooses the inexpensive dish. However, for a single round of the Diners Dilemma, each individual is better off choosing the expensive dish no matter what the others order, and thus Nash equilibrium is achieved when everyone chooses the expensive dish. When the same group of diners meets repeatedly under the same bill-sharing agreement, tacit cooperation may develop, leading to a better overall group dining experience in this simplified scenario.

To our knowledge, no investigations exist regarding SA in games of strategic interaction like the Diner's Dilemma. Intuitively, SA and the different levels of SA should play an important role in the levels of cooperation in this game, and as a consequence, in the levels of trust. Related investigations regarding the effects of interdependency information and the availability of such information to players of the Iterated Prisoner's Dilemma seem to suggest that the information provided would highly influence the level of cooperation [15], [9]. Information about each others outcomes and actions becomes especially important for decision-makers to infer one anothers intentions and to predict one anothers actions. For example, Rapoport and Chammah (1965) [15] demonstrated that participants who viewed a payoff matrix in the Prisoner's Dilemma cooperated more than those who earned payoffs through experience alone (46% versus 22% of actions). Unlike individual decision-making, these uniquely social factors make it possible for individuals to alter each other' s behavior [9].

Gonzalez and colleagues have tested the effect of different levels of information in the proportion of cooperation in the Prisoner's Dilemma. In a laboratory experiment, pairs of participants were given differing levels of interdependence information across four conditions: No-Info players saw only their own actions and outcomes, and were not told that they interacted with another person; Min-Info players knew they interacted with another person, but still without seeing the others actions or outcomes; Mid-Info players discovered the others actions and outcomes as they were revealed over time; and Max-Info players were also shown a complete payoff matrix mapping actions to outcomes throughout the game. Except for similar behavior in the No-Info and Min-Info conditions, additional interdependence information increased individual cooperation and mutual cooperation, driven by increased reciprocated cooperation (in response to a counterparts cooperation). Furthermore, joint performance and satisfaction were generally higher for pairs with more information.

The findings reviewed above clearly indicate that awareness of interdependence may encourage pro-social behavior and trust in many real-world interactions. Research on interactive user interfaces for collaborative decision-making highlights the importance of revealing trust-influencing information at different levels of granularity [18], [2]. We investigate the effect of three levels of awareness as guided by the information presented at the interface: SA Levels 1, 2, and 3 in the Diner' s Dilemma. Intuitively, we expect that interfaces that provide information of higher SA Levels will encourage pro-social behavior when the interface visually suggests that opponents are potentially cooperative and will encourage defection when opponents are consistently exploitive with no chance of conversion to cooperation.

III. APPROACH

The focus and novel contribution of this paper is the model-based assessment of trust and SA in the Diner's Dilemma game. To provide readers a clear understanding of the data

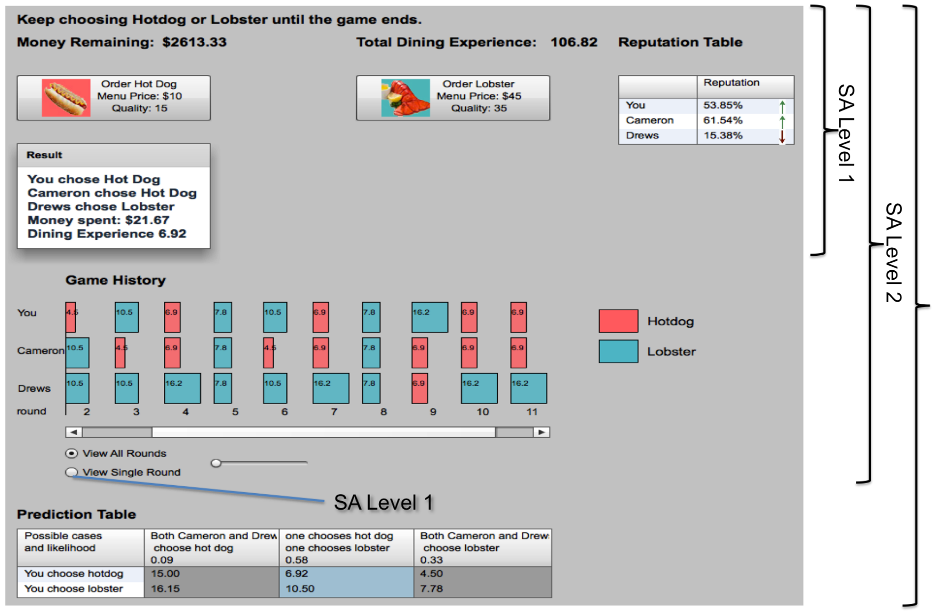


Fig. 1. Splash Screen of the Diner's Dilemma Browser-based Game (from Teng et al. [19]). The scale on the right side shows the displayed components for each SA Level. For example, SA 1 does not show the Prediction Tables. One exception to this visual categorization is that at SA Level 1, participants see the Game History component in "single round" mode only.

used to build these models, here we briefly revisit the implementation and experimental setup from Teng's study in [19].

A. Visualizing SA Levels in the Diner's Dilemma Game

Figure 1 shows an annotated screenshot of the browser-based implementation of the Iterative Diner's Dilemma game. The legend on the right side shows the components that are displayed to participants at each SA Level. SA Level 1 (Perception) has the most basic component set, consisting of price and quality information for both hot dog and lobster, and current money remaining and points gained. SA Level 1 also shows a reputation panel which provides some insight into the overall reputation of each player based on their history of actions in the game. The reputation panel shows the deviation of each player's reputation from that of the previous round, denoted by a red or green directional arrow. SA Level 1 also shows the game history component, but only in "single round" mode.

SA Level 2 (Comprehension) adds more data to the game history horizontal bar graph that allows participants to view the actions of each opponent over the previous rounds, to promote a better understanding the strategies of the opponent players. Players have the option to view all rounds or a single round. SA Level 3 (Projection) adds a table that projects the likelihood of various different outcomes based on the current game history.

B. Opponent Strategies

To explore the deviations in trust behavior at different SA levels, it was necessary to employ a variety of different opponent strategies, spanning from cooperation-encouraging defection-encouraging, shown in the two lists below, respectively.

- *Stimulate*: If the participants cooperation percentage so far in the current block of the game is above 66% and the participant chose hotdog in the last round, both computer agents will choose hotdog in the next round; case b): if its below 33% and the participant chose lobster in the last round, both computer agents will choose lobster; case c): otherwise one agent will choose hotdog and the other will choose lobster.
- *StimulateNoise*: Similar to Stimulate, with the addition of noise. In cases a) and b), the computer agents decisions are reversed from what they would have been in Stimulate for 10% of the rounds.
- *Tricky*: For every other round (0,2,4,6...), both opponents will play "Stimulate" (and only consider the players every other rounds decision as well). For the remainder (1, 3, 5, 7...), they will play a random move.
- *Negative30*: In this strategy, the combined cooperation of the two opponents is always 70%. Initially, Opponent 1 will play Tit-for-Tat against the participant and Opponent 2 will always make the decision which makes the combined cooperation of the two opponents closest to 70%. Each of the two opponents makes 50 decisions in the course of a block, for a combined total of 100 decisions. If the two opponents reach a combined total of 70 cooperate (hot dog) decisions (70% of 100 decisions) before the end of the block, both opponents will choose to defect for the remaining rounds.
- *Negative70*: Same as Negative 30, except that the combined cooperation is 30%.

- *Random*: The two opponents randomly choose between the two meals with equal probability in each round.

C. Self-Reported Trust

As we have described earlier, our model-based evaluation focuses on two aspects of trust. *Observed* trust is simply recorded as the degree of cooperation in each condition of the game. To gather information on the perceived trust in opponent players throughout the game, and to garner additional subjective information about participants mental models of the game, survey questions were administered between trials in the game. Figure 2 shows a screen shot of an example questionnaire form for SA Level 3. These questions are described in more detail in [19]. Data from these self-assessments are used to build subjective models of trust, and we show in Section III-C that they can serve as a strong predictor of observed trust in all conditions of the game. In our evaluations, we first examine the predictive power of self-reported data and condition data (SA Level and Strategy) independently, since the self-reported data contains specific information about individual participants' opinions during the game. The separation allows us to assess predictive ability for trust behavior in the absence of self-reported data.

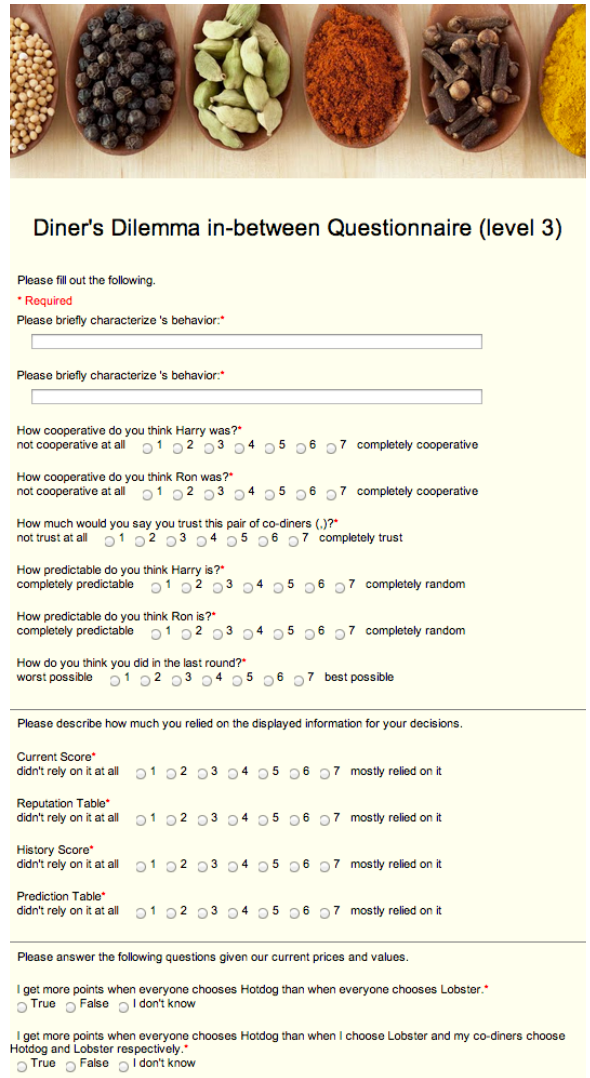
IV. Analysis and Results

Teng et. al [19] presented a correlation based analysis showing that multiple factors influence the relation between trust and situational awareness in the context of the 3-Player Diners Dilemma game. We now examine multiple influence factors in parallel from a model-based learning perspective. Results from the 24 studies were loaded into the WEKA machine learning toolkit[1] to assess the performance of a variety of model and rule-based learners over the collected data. For all models discussed in this section, a train-test split of 60:40 was applied and 10-fold cross validation leave-one-out analysis was used to assess predictive capacity of each. That is, how well each model could predict the degree of trust (cooperation) behavior for each condition. As a preprocessing step, 144 sets of trust behavior data were binned into two classes, Low, High based on an equal distribution around the mean of 47.3%. Since we are interested in both observed and self-reported trust information, we first train a set of machine learning algorithms to predict observed trust behavior using only Survey (self-reported) data. Next, we attempt to predict it using models with only SA Level, Score and Opponent Strategy data. Last, we combine all available data and attempt to build a prediction model for observed trust. This includes a discussion of a best-first feature analysis, describing the predictive ability of each feature used.

A. Distribution and Correlation Analysis

Before we present our discussion of the various classification algorithms on each data set, we now present a brief overview of distribution and correlations of the salient features in the study. [19] provides a more detailed statistical analysis

[1]http://www.cs.waikato.ac.nz/ml/weka/



Diner's Dilemma in-between Questionnaire (level 3)

Please fill out the following.

* Required

Please briefly characterize 's behavior:*

Please briefly characterize 's behavior:*

How cooperative do you think Harry was?*
not cooperative at all 1 2 3 4 5 6 7 completely cooperative

How cooperative do you think Ron was?*
not cooperative at all 1 2 3 4 5 6 7 completely cooperative

How much would you say you trust this pair of co-diners (,)?*
not trust at all 1 2 3 4 5 6 7 completely trust

How predictable do you think Harry is?*
completely predictable 1 2 3 4 5 6 7 completely random

How predictable do you think Ron is?*
completely predictable 1 2 3 4 5 6 7 completely random

How do you think you did in the last round?*
worst possible 1 2 3 4 5 6 7 best possible

Please describe how much you relied on the displayed information for your decisions.

Current Score*
didn't rely on it at all 1 2 3 4 5 6 7 mostly relied on it

Reputation Table*
didn't rely on it at all 1 2 3 4 5 6 7 mostly relied on it

History Score*
didn't rely on it at all 1 2 3 4 5 6 7 mostly relied on it

Prediction Table*
didn't rely on it at all 1 2 3 4 5 6 7 mostly relied on it

Please answer the following questions given our current prices and values.

I get more points when everyone chooses Hotdog than when everyone chooses Lobster.*
True False I don't know

I get more points when everyone chooses Hotdog than when I choose Lobster and my co-diners choose Hotdog and Lobster respectively.*
True False I don't know

Fig. 2. Screen shot of a sample questionnaire showing a selection of the Self-reported trust questions from the Diner's Dilemma study.

Fig. 3. Distribution of participant responses to the self-reported survey questions, partitioned by their observed trust behavior for each instance. (Blue = Low, Red = High)

of correlations in the Diner's Dilemma study data set. Figure 3 shows a set of scatterplots depicting pairwise correlation between our two independent variables (Opponent Strategy and S.A. Level) and our dependent variables (Binned Trust and Total Score). A small amount of jitter has been added to the plots to highlight sizes of the various clusters in the data. Interestingly, Figure 3 shows slightly increased observed trust behavior at low SA (indicated by the larger red cluster in the top left corner, relative to the blue beneath it). In the next plot to the right, observed trust behavior appeared evenly distributed across strategies. Next, high trust behavior tended to produce a broader distribution of overall scores, indicating that trust behavior had a win-big or lose-big effect, relative to conservative, defecting behavior. The rightmost plot shows that there are fairly equal numbers of data points in the trust bins. The second row shows us that in general, higher scores were achieved by trusting participants, most notably at SA Level 1, as shown in the Score:SA Level plot. The strategy plots indicate that there is an effect of strategy on total score, but trust behavior appears to be evenly distributed across strategies.

a) Predicting Observed Trust Behavior Using Survey Data: Sandholm et al. [16] describe a study of various machine learning algorithms playing the iterated prisoner's dilemma game. They find that prediction algorithms with greater historical information, lookup table memory, and long exploration schedules (for reinforcement learning) tend to perform best. In this study our evaluation focuses on model-based learners only. In contrast to reinforcement learners, wherein a model is constantly refined at each step as a game is played, the machine learning algorithms tested here learn different types of prediction models based on a training set of data, and we then evaluate the accuracy of the models on a separate test data set. In short, model-based learners apply a static prediction model, relative to reinforcement learners. More recent work by Caruana and Niculescu-Mizil [3] examines the general performance of supervised learning algorithms in an empirical study. In particular, they evaluate Support Vector Machines, neural nets, logistic regression, naive bayes, memory-based learning, random forests, decision trees, bagged trees, boosted trees, and boosted stumps on eleven binary classification problems using a variety of performance metrics. [3] concludes that the RandomForest algorithms produced excellent results and all tasks. They also report that the Naive Bayes algorithm showed poorer performance on average. However, the key finding from [3] was that a significant variability across the problems and metrics tested. Based on these findings, we set an initial expectation that a RandomForest algorithm will perform better than other classification algorithms on our Diner's Dilemma data, but we also choose a diverse selection of learning algorithms because of the high variance reported in [3].

In total, 6 machine learning algorithms were trained and tested on each data set ("self reported data", "SA Level and strategy only", and "all data"), and predictive accuracy was analyzed for each. The algorithms, listed in Table I, include

Algorithm	TP	FP	F	ROC	% Accuracy
Survey Data Only					
Zero-r	0.535	0.535	0.373	0.47	0.53
K-Star	0.653	0.339	0.652	0.705	65.27
Naive Bayes	0.708	0.293	0.709	0.762	70.83
Rand. Forest	0.667	0.342	0.665	0.766	66.66
J-48	0.688	0.316	0.687	0.712	68.75
Avg-All	0.6502	0.365	0.6172	0.683	64.986
SA Level and Strategy Data Only					
Zero-r	0.535	0.535	0.373	0.47	0.53
K-Star	0.632	0.378	0.63	0.601	63.19
Naive Bayes	0.597	0.416	0.593	0.652	59.72
Rand. Forest	0.59	0.419	0.589	0.629	59.02
J-48	0.674	0.348	0.665	0.591	67.36
Avg-All	0.6056	0.4192	0.57	0.5886	60.542
All Data					
Zero-r	0.535	0.535	0.373	0.47	0.53
K-Star	0.708	0.293	0.709	0.77	70.83
Naive Bayes	0.667	0.333	0.667	0.732	66.66
Rand. Forest	0.736	0.265	0.736	0.825	73.36
J-48	0.764	0.24	0.764	0.791	76.38
Avg-All	0.682	0.3332	0.6498	0.7176	68.13

TABLE I

RESULTS OF DIFFERENT CLASSIFICATION ALGORITHMS PREDICTING TWO NOMINAL CLASSES OF OBSERVED TRUST BEHAVIOR USING DIFFERENT DATA SETS COLLECTED FROM THE DINER'S DILEMMA STUDY.

a simple majority class predictor (Zero-r); a Naive Bayes probabilistic learner; a K-Star lazy learner; Random Forest and J-48 tree-based learners, and a simple averaging strategy over all of these. The same set of algorithms were applied to each data set, and comparative performance results are shown in Table I.

To explore the relationship between observed trust behavior and self reported trust in various aspects of the system (Figure 2), a set of prediction models were trained based only on the subjective (perception) data from the self-reported questionnaire data. Table I shows the model that best predicted trust-bins (High, Low) on the subjective data was the Naive Bayes method, with an accuracy of 70.83%. This is promising result because it indicates that perception about the situation (perception of independent variables such as opponent strategy) do influence participant behavior. This result is a relative improvement of 41% over the simple majority class predictor, which scored only 53%. Table I also shows true and false positive rates, F-measuree and Receiver Operator Characteristic for each method.

b) Predicting Trust Behavior Using SA Level and Strategy Data : Next, we isolate our independent variables (SA Level and Opponent Strategy) and use them as predictors of observed trust within the Diner's Dilemma game. Results are shown in the middle section of Table I for each of the 6 prediction algorithms. Interestingly, the algorithms perform quite differently, and are approximately 4% less accurate on average at predicting observed trust. The best performing algorithm on this data is clearly the J48 rule-based learner, which correctly predicted 67% of trust decisions (compared with 53% for the simple majority predictor. Figure 5 shows a comparative analysis of the accuracy scores for each algorithm.

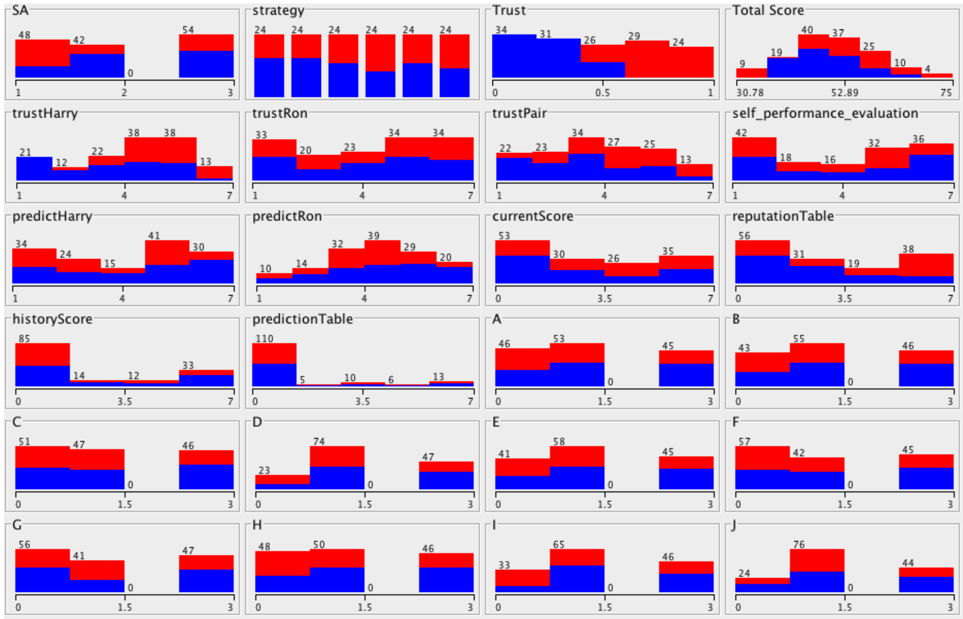


Fig. 4. Distribution of participant responses to the self-reported survey questions, partitioned by their observed trust behavior for each instance.

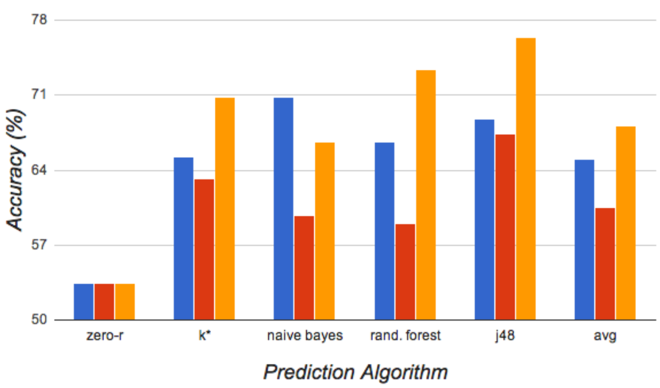


Fig. 5. Predictive accuracy results for 6 different learning algorithms, grouped by data set. Blue columns on the left represent Survey data only, red columns (middle) represent SA Level and Strategy data only, and the orange (right) columns represent a combination of all features.

Each algorithm has three associated columns, representing self-reported, SA Level / Strategy, and all data, from left to right.

c) Predicting Trust Behavior Using All Data: Now that we have analyzed the predictive capacity of self-reported data and condition-based data on trust behavior, a logical next step is to combine both into a joint model. This step in the analysis revealed our most positive result, which is a clear synergy between self-reported data and condition-based data for prediction of trust behavior in the game. All of the learning algorithms that operated on the joint data set significantly outperformed previous results. The overall best performer was again the J48 learner with an accuracy of 76.3 % (details in Table I. However, the biggest improvement over the other data sets was exhibited by the Random Forest algorithm, which

```
J48 pruned tree
------------------

trustHarry <= 2: low (21.0)
trustHarry > 2
|   reputationTable <= 5
|   |   trustHarry <= 4
|   |   |   historyScore <= 0
|   |   |   |   self_performance_evaluation <= 3: high (4.0/1.0)
|   |   |   |   self_performance_evaluation > 3: low (2.0)
|   |   |   historyScore > 0: low (18.0)
|   |   trustHarry > 4
|   |   |   trustHarry <= 6
|   |   |   |   predictRon <= 6
|   |   |   |   |   C <= 0: high (10.0/1.0)
|   |   |   |   |   C > 0
|   |   |   |   |   |   J <= 1
|   |   |   |   |   |   |   I <= 0: high (4.0/1.0)
|   |   |   |   |   |   |   I > 0
|   |   |   |   |   |   |   |   G <= 0
|   |   |   |   |   |   |   |   |   predictRon <= 5
|   |   |   |   |   |   |   |   |   |   reputationTable <= 4: high (3.0)
|   |   |   |   |   |   |   |   |   |   reputationTable > 4: low (3.0)
|   |   |   |   |   |   |   |   |   predictRon > 5: low (4.0)
|   |   |   |   |   |   |   |   G > 0: low (4.0)
|   |   |   |   |   |   J > 1
|   |   |   |   |   |   |   trustPair <= 4
|   |   |   |   |   |   |   |   predictRon <= 4
|   |   |   |   |   |   |   |   |   trustHarry <= 5: high (2.0)
|   |   |   |   |   |   |   |   |   trustHarry > 5: low (3.0/1.0)
|   |   |   |   |   |   |   |   predictRon > 4: low (3.0)
|   |   |   |   |   |   |   trustPair > 4: high (8.0/1.0)
|   |   |   |   predictRon > 6: low (7.0)
|   |   |   trustHarry > 6: high (11.0/1.0)
|   reputationTable > 5
|   |   historyScore <= 2: high (17.0)
|   |   historyScore > 2
|   |   |   trustRon <= 4: high (8.0/1.0)
|   |   |   trustRon > 4
|   |   |   |   trustRon <= 5: low (8.0/1.0)
|   |   |   |   trustRon > 5: high (4.0)
```

Fig. 6. Example of a J-48 Decision Tree learned from the survey (self-reported) assessments. The goal of the algorithm was to predict observed trust behavior in a hidden test set.

scored 73.3%, or a 23% improvement over the condition data alone, and an improvement of 15% over the self-reported data alone.

B. Best-Feature Analysis

To determine which particular features were the best indicators of trust behavior, a best-first analysis was performed over all features. For the subjective variables (self-reported data), perceived trust in opponent 1 was the top predictor of trust, followed by perceived importance of the reputation table. For the objective variables, SA Level and Strategy were the top predictors of trust behavior respectively. In fact, participant score was the strongest predictor overall, but this was omitted from all of our model-based evaluations, since it represents some after-the-fact knowledge of participant behavior, which contradicts the purpose of the prediction models.

V. Conclusion

This paper has presented a model-based approach to learning the relationship between trust behavior (observed and self-reported) and situational awareness in the context of the 3 person Iterative Diner's Dilemma game. Our analysis used data recorded from a study reported in [19]. A set of 6 machine learning algorithms were applied to predict observed trust based on three underlying data slices: self-reported data, SA Level and Opponent strategy data, and a combination of both. A follow-up study will focus on the performance of case-based and instance-based learners [1] against model-based learners for the purpose of predicting human behavioral data in trust-games. This paper has shown that the best overall predictor of observed trust in terms of predictive accuracy was a J48 rule-based learner, trained on the combination data set. However, for the self-reported data, the J48 learner was outperformed by a simple Naive Bayes learner. All of the learning approaches exhibited at least 10% improvement over a baseline majority class predictor, indicating that both self reported trust, and SA Level / Strategy data are good predictors of observed trust behavior in the Diner's Dilemma game.

Acknowledgment

Manuscript received October 25, 2012. This material is based in part upon work supported by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053 and by NSF grant IIS-1058132. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory, NSF, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes not-with-standing any copyright notation here on.

References

[1] David W. Aha and Dennis Kibler. Instance-based learning algorithms. In *Machine Learning*, pages 37–66, 1991.

[2] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. Tasteweights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*, RecSys '12, pages 35–42, New York, NY, USA, 2012. ACM.

[3] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning*, ICML '06, pages 161–168, New York, NY, USA, 2006. ACM.

[4] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. Information credibility on twitter. In *WWW*, pages 675–684, 2011.

[5] M. R. Endsley. Situation awareness global assessment technique (SAGAT). pages 789–795 vol.3, May 1988.

[6] M. R. Endsley. A survey of situation awareness requirements in air-to-air combat fighters. *The International Journal of Aviation Psychology*, 3(2):157–168, 1993.

[7] M. R. Endsley. Toward a theory of situation awareness in dynamic systems: Situation awareness. *Human factors*, 37(1):32–64, 1995.

[8] J. Golbeck and J. Hendler. Inferring binary trust relationships in web-based social networks. *ACM Transactions on Internet Technology (TOIT)*, 6(4):497–529, 2006.

[9] Cleotilde Gonzalez and Jolie M. Martin. Scaling up instance-based learning theory to account for social interactions. *Negotiation and Conflict Management Research*, 4(2):110–128, 2011.

[10] Rashaad Jones, Erik Connors, Mary Mossey, John Hyatt, Neil Hansen, and Mica Endsley. Using fuzzy cognitive mapping techniques to model situation awareness for army infantry platoon leaders. *Computational & Mathematical Organization Theory*, 17:272–295, 2011. 10.1007/s10588-011-9094-6.

[11] Rashaad E. Jones, Erik S. Connors, Mary E. Mossey, John R. Hyatt, Neil J. Hansen, and Mica R. Endsley. Using fuzzy cognitive mapping techniques to model situation awareness for army infantry platoon leaders. *Comput. Math. Organ. Theory*, 17(3):272–295, September 2011.

[12] Byungkyu Kang, John O'Donovan, and Tobias Höllerer. Modeling topic specific credibility on twitter. In *IUI*, pages 179–188, 2012.

[13] J. O'Donovan, B. Kang, G. Meyer, T. Hllerer, and S. Adali. Credibility in context: An analysis of feature distributions in twitter. In *In proceedings at the 2012 ASE/IEEE International Conference on Social Computing, SocialCom, Amsterdam, The Netherlands.*, pages 167–174. IEEE, 2012.

[14] J. O'Donovan and B. Smyth. Trust in recommender systems. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pages 167–174. ACM, 2005.

[15] A. Rapoport and A.M. Chammah. *Prisoner's dilemma*, volume 165. University of Michigan Press, 1965.

[16] Tuomas W. Sandholm and Robert H. Crites. Multiagent reinforcement learning in the iterated prisoner's dilemma. *Biosystems*, 37:147–166, 1995.

[17] L.D. Strater, D. Jones, and M.R. Endsley. Analysis of infantry situation awareness training requirements. Technical report, DTIC Document, 2001.

[18] Kirsten Swearingen and Rashmi Sinha. Beyond algorithms: An hci perspective on recommender systems, 2001.

[19] Y. Teng, R. Jones, J. O'Donovan, L. Marusich, Gonzalez. C., and T. Hollerer. Trust and situation awareness in a 3-player diner's dilemma game (currently under review). In *2012 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support*. IEEE.