

Agent Assistants for Team Analysis

Milind Tambe, Taylor Raines, and Stacy Marsella

■ With the growing importance of multiagent teamwork, tools that can help humans analyze, evaluate, and understand team behaviors are also becoming increasingly important. To this end, we are creating isaac, a team analyst agent for post hoc, offline agent-team analysis. ISAAC's novelty stems from a key design constraint that arises in team analysis: Multiple types of models of team behavior are necessary to analyze different granularities of team events, including agent actions, interactions, and global performance. These heterogeneous team models are automatically acquired by machine learning over teams' external behavior traces, where the specific learning techniques are tailored to the particular model learned. Additionally, ISAAC uses multiple presentation techniques that can aid human understanding of the analyses. This article presents ISAAC's general conceptual framework and its application in the RoboCup soccer domain, where ISAAC was awarded the RoboCup Scientific Challenge Award.

Multiagent teamwork is an important area of agent research, with a growing number of applications, including multirobotic space missions, virtual environments for training and education, and software agents on the internet. With the growing importance of teamwork, there is now a critical need for tools to help humans analyze, evaluate, and understand team behaviors. Indeed, in multiagent domains with tens or even hundreds of agents in teams, agent interactions are often highly complex and dynamic, making it difficult for human developers to analyze agent-team behaviors. The problem is further exacerbated in environments where agents are developed by different developers, where even the intended interactions are unpredictable.

Unfortunately, the problem of analyzing team behavior to aid human developers in understanding and improving team performance has largely been unaddressed. Previous work in agent teamwork has largely focused on guiding agents in teamwork (Tambe 1997) but

not on its analysis for humans. Agent explanation systems, such as DEBRIEF (Johnson 1994), allow individual agents to explain their actions based on internal state but do not have the means for a team analysis. Recent work on multiagent visualization systems, such as Nduму et al. (1999), has been motivated by multiagent understandability concerns (similar to ours), but it still leaves analysis of agent actions and interactions to humans.

This article focuses on agents that assist humans to analyze, understand, and improve multiagent team behaviors by (1) locating key aspects of team behaviors that are critical in team success or failure; (2) diagnosing such team behaviors, particularly problematic behaviors; (3) suggesting alternative courses of action; and (4) presenting the relevant information to the user comprehensibly. To accomplish these goals, we have developed an agent called ISAAC. A fundamental design constraint here is that unlike systems that focus on explaining individual agent behaviors (Johnson 1994), team analysts such as ISAAC cannot focus on any single agent or any single perspective or any single granularity (in terms of time scales). Instead, when analyzing teams, multiple perspectives at multiple levels of granularity are important. Thus, although it is sometimes beneficial to analyze the critical actions of single individuals, at other times, it is the collaborative agent interaction that is key in team success or failure and requires analysis, yet at other times, an analysis of the global behavior trends of the entire team is important.

To enable analysis from such multiple perspectives, ISAAC relies on multiple models of team behavior, each covering a different level of granularity of team behavior. More specifically, ISAAC relies on three heterogeneous models that analyze events at three separate levels of granularity: (1) an individual agent action, (2) agent interactions, and (3) overall team

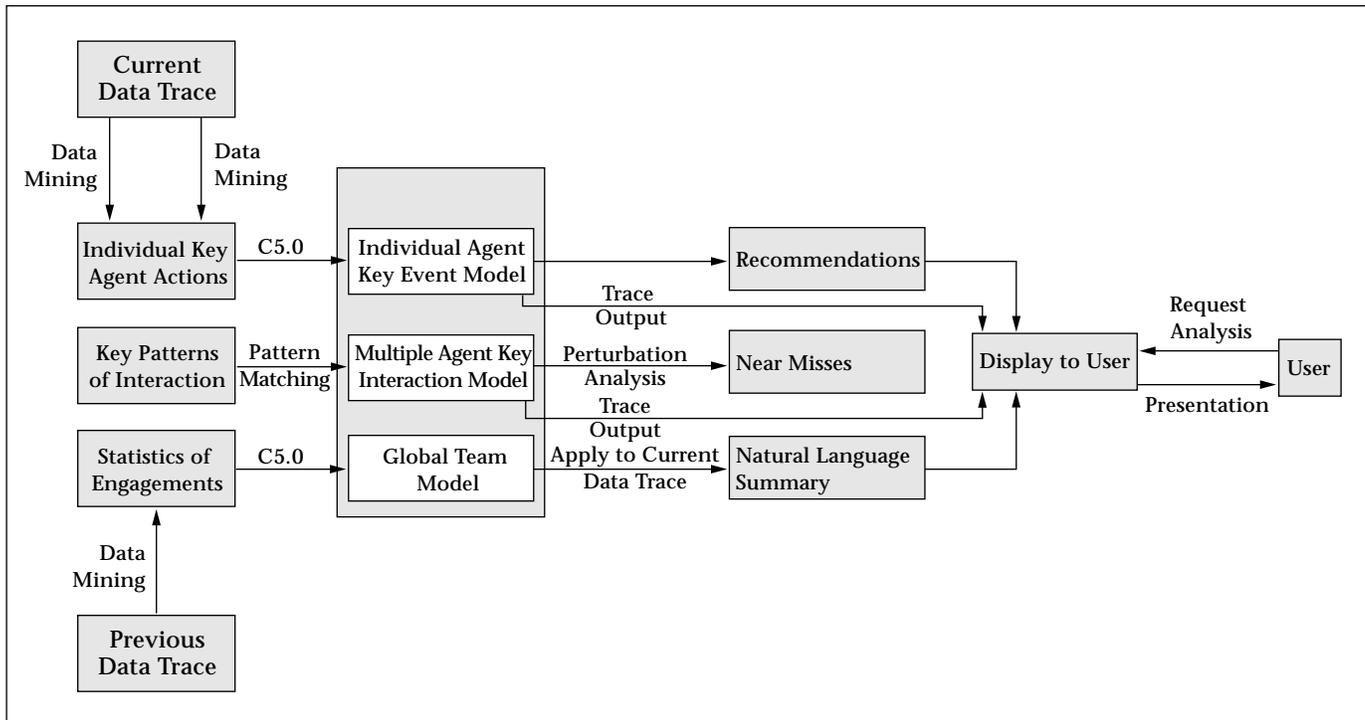


Figure 1. Flowchart for ISAAC Model Generation and Analysis.

behavior. These models are automatically acquired using different methods (inductive learning and pattern matching); indeed, with multiple models, the method of acquisition can be tailored to the model being acquired.

The constraint of multiple models has strong implications for the type of presentation to humans as well. Analysis of an agent action can show the action and highlight features of the action that played a prominent role in its success or failure, but a similar presentation would be incongruous for a global analysis because no single action would suffice. Global analysis requires a more comprehensive explanation that ties together seemingly unconnected aspects and trends of team behavior. ISAAC utilizes a natural language summary to explain the team's overall performance, utilizing its multimedia viewer to show examples where appropriate. The content for the summary is chosen based on ISAAC's analysis of key factors determining the outcome of the engagement. Additionally, ISAAC presents alternative courses of action to improve a team using a technique called *perturbation analysis*. A key feature of perturbation analysis is that it finds actions within the agents' skill set, such that recommendations are plausible.

ISAAC is currently applied in the RoboCup soccer simulation domain, an environment for multiagent research. Here, 2 teams of 11 agents each play soccer in a noisy, dynamic environ-

ment. Agent-team analysis is critical in RoboCup because team developers want to understand the strengths and weaknesses of teams and improve such teams. Indeed, ISAAC has been applied to all the teams from several RoboCup tournaments, revealing many interesting results, including surprising weaknesses of leading teams in previous RoboCup tournaments. ISAAC won the Scientific Challenge Award at the RoboCup-99 international tournament.¹

Analysis Using Multiple Models

ISAAC's analysis can be split into two phases: (1) model acquisition and (2) model utilization. An overview of the entire process is shown in figure 1. For model acquisition, input to all models comes from data traces of agent behaviors where the traces are uploaded from users around the world through the internet. By using data from the agents' external behavior traces, ISAAC is able to analyze a team without necessarily understanding its internals, allowing analysis of teams developed by different developers. ISAAC applies inductive learning and pattern-matching algorithms to the traces to acquire its heterogeneous models. In particular, analysis of an individual agent action (*individual agent key event model*) uses the c5.0 decision tree inductive-learning algorithm, an

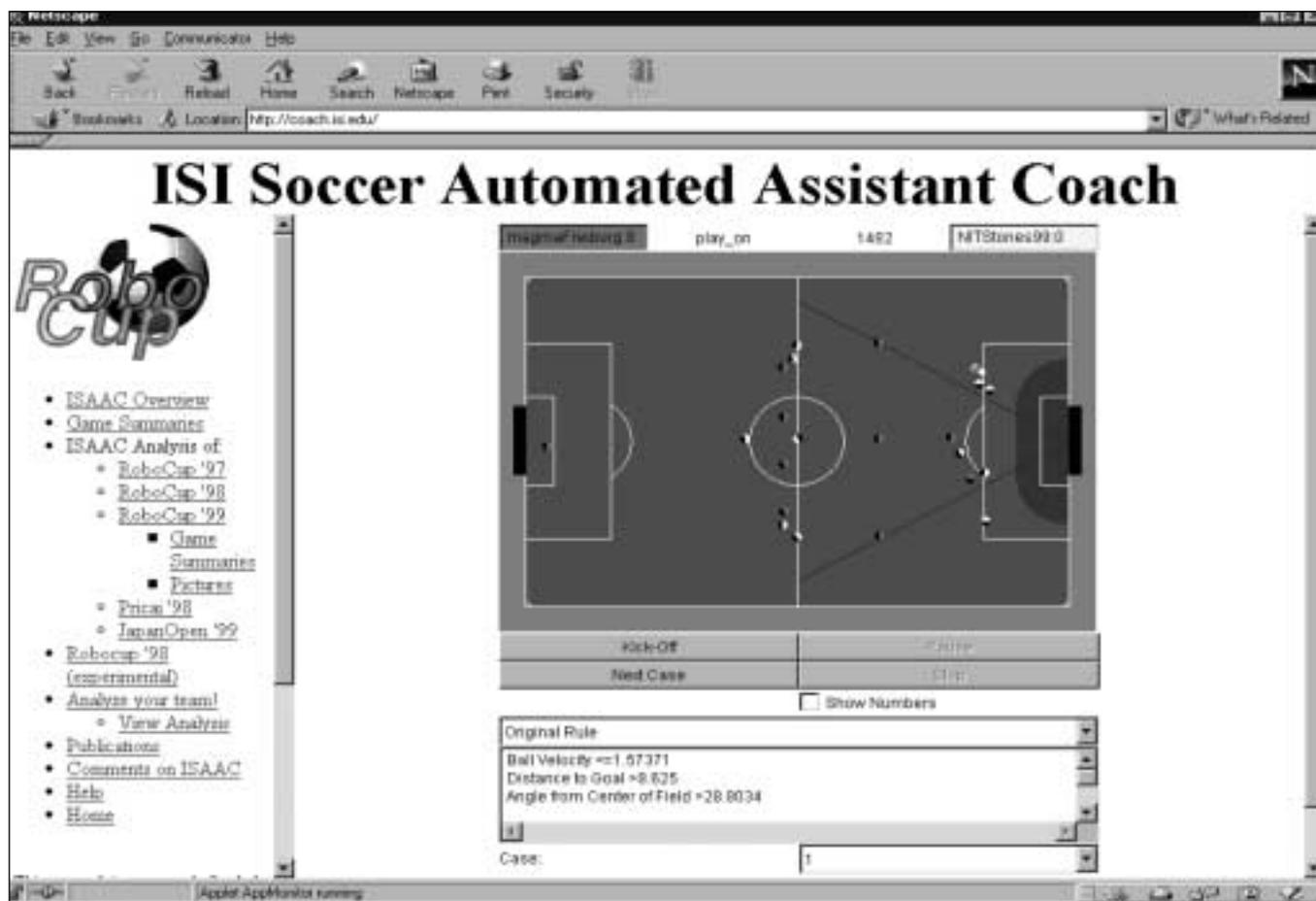


Figure 2. ISAAC's Multimedia Viewer.

extension to c4.5, to create rules of success or failure (Quinlan 1994). For analysis of agent interactions (*multiple agent key interaction model*), predefined patterns are matched to find prevalent patterns of success. To develop rules of team successes or failures (*global team model*), game-level statistics are mined from all available previous games, and again, inductive learning is used to learn rules that determine game success and failure.

Using the models involves utilizing a different presentation technique at each granularity of analysis to maximize human understandability. For the individual agent key event model, the rules and the cases they govern are displayed to the user. By themselves, the features that compose a rule provide implicit advice for improving the team. To enable the user to further understand the situation and validate the rules, a multimedia viewer is used to show cases matching the rule (figure 2). A perturbation analysis is then performed to recommend changes to the team by changing the rule con-

dition by condition and mining cases of success and failure for this perturbed rule. The cases of this analysis are also displayed in the multimedia viewer.

For the multiple agent key interaction model, patterns of agent actions are analyzed similar to the individual agent actions. A perturbation analysis is also performed here to find patterns that are similar to successful patterns but were unsuccessful. Both successful patterns and these "near misses" are displayed to the user as implicit advice.

The global team model requires a different method of presentation. Here, the current engagement is matched against previously learned rules. ISAAC considers any matching rule(s) as providing the reasons for the outcome of the current engagement. A natural language summary of the engagement is generated using this rule for content selection and sentence planning. ISAAC uses the multimedia display here as well, linking text in the summary to corresponding selected highlights. The

Haarlem Offense Collapses in Stunning Defeat at the Hands of 11MONKEYS!

11MONKEYS displayed their offensive and defensive prowess, shutting out their opponents 7-0. 11MONKEYS pressed the attack hard against the HAARLEM defense, keeping the ball in their half of the field for 84% of the game and allowing ample scoring opportunities. HAARLEM pulled their defenders back to stop the onslaught but to no avail. To that effect, 11MONKEYS was able to get past HAARLEM's last defender, creating two situations where only the goalie was left to defend the net. 11MONKEYS also handled the ball better, keeping control of the ball for 86% of the game. HAARLEM had a tendency to keep the ball toward the center of the field as well, which might have helped lead them to ruin given the ferocity of the 11MONKEYS attack.

sidebar presents an example of a natural language summary produced by ISAAC, where ISAAC emphasizes the reasons why 11MONKEYS defeated the HAARLEM team. The underlined sentences correspond directly to the matching rule (the title and first sentence are generated separately based on headlines taken from human soccer world cup games).

Evaluation and Results

To evaluate ISAAC, we first evaluate each of its models in isolation and then the effectiveness of the integrated ISAAC system. We begin by evaluating the individual agent model. A key measure of this model is the effectiveness of its analysis, specifically the capability to discover novel patterns. Here, ISAAC has been able to find some surprises in the top RoboCup teams. For example, ISAAC found a problematic pattern in the shooting behavior of ANDHILL'97, the second-place winner in 1997. Not only was this problematic pattern surprising to us, but it was also surprising to the developer of the team, Tomohito Andou. After hearing of this result, and witnessing it through ISAAC's multimedia interface, he told us that he "was surprised that ANDHILL's goal shooting behavior was so poor..." and "... this result would help improve [the] ANDHILL team in the future."²

Another interesting result from the individ-

ual agent analysis model comes from the number of rules governing shooting behavior and defensive prowess. ISAAC's analysis shows that the number of rules for defense decreased for the top four teams (from nine rules to five rules), perhaps indicating more refined defensive structures as the teams progress. Also, the number of rules necessary to capture the behavior of a team's offense is consistently more than that necessary for defense (about 10 to 15 rules), possibly because no single offensive rule could be effective against all opponent defenses. The key here is that global analysis of team behaviors is now within reach with team-analyst tools such as ISAAC.

Another point of evaluation is how well ISAAC models the shooting behaviors. To this end, ISAAC models were applied to predict game scores at RoboCup-99. ISAAC used rules describing a team's defense and matched them with the raw averaged data of the shots taken by the other team to produce an estimate of how many goals would be scored against the team in the upcoming game. Performing this analysis for both teams produced a predictive score for the outcome of the game. This prediction obviously ignores many critical factors; for example, some early games were unrepresentative, and some teams were changed by hand during the competition. However, in practice, ISAAC's predictive accuracy was 70 percent with respect to wins and losses, indicating it had managed to capture the teams' defenses quite well in its model.

For game summaries, one measure is a comparison of the number of features used in the current summaries versus those generated earlier that did not use ISAAC's approach. On average, ISAAC uses only about 4 features from its set of 10 statistics in the summaries, resulting in a 60-percent reduction from a natural language generator not based on ISAAC's machine-learning-based analysis. Thus, ISAAC's approach was highly selective in terms of content. Indeed, summaries generated without ISAAC were much longer, lacked variety, and failed to emphasize the key aspects of the game. The audience also appeared to approve of ISAAC's summaries. In a small survey conducted at RoboCup-99 (about 20 participants), 75 percent of the participants rated the summaries as *very good*, with another 10 percent rating the summaries as *good*.

Yet another measure of ISAAC's use of the team model for natural language generation is available by viewing the error rates from the machine-learning algorithm used. These error rates tell us how accurately ISAAC's learned rules reflected the game. On the original set of games for which ISAAC's rules were learned, 87 percent

of the games were classified correctly, resulting in an error rate of 13 percent. Our test set of (unseen) RoboCup-99 games resulted in 72 percent being classified correctly, for an error rate of 28 percent. If an error does occur, ISAAC still produces a summary, but it reflects its surprise at the outcome, thus explaining the error. The high error rate on our training data could indicate that a better feature set is possible or that the data might be noisy.

Evaluating ISAAC as an integrated system is more difficult. However, some observations can still be made. ISAAC was awarded the Scientific Challenge Award at RoboCup-99 in Stockholm. At RoboCup-99, developers used ISAAC to analyze opponent teams after the early round matches to get a feel for the skill of upcoming opponents. Also, spectators and developers alike were able to view ISAAC's summaries just minutes after a game.

Conclusion

With the growing importance of multiagent teamwork, it is now increasingly critical to build automated assistants to aid developers in analyzing agent-team behaviors. We have taken a step toward this goal by building an agent called ISAAC for post hoc, offline agent-team analysis. ISAAC uses two key novel ideas in its analysis: First, ISAAC utilizes multiple models of team behavior to analyze different granularities of agent actions, using inductive learning techniques, enabling the analysis of differing aspects of team behavior. Second, ISAAC supports perturbations of models, enabling users to engage in what-if reasoning about the agents and providing suggestions to improve agents' performance. Additionally, ISAAC combines multiple presentation techniques to aid humans in understanding the analysis, where presentation techniques are tailored to the model at hand. Although ISAAC has been applied in the context of the RoboCup soccer simulation, we hope to apply it to other domains, such as battlefield simulations (Tambe 1997), in the future.

Acknowledgments

We thank Intel Corporation for its generous support of the work reported in this article.

Notes

1. ISAAC is available on the web at coach.isi.edu.
2. Personal communication with Tomohito Andou.

References

Johnson, W. L. 1994. Agents That Learn to Explain Themselves. In Proceedings of the Twelfth National Conference on Artificial Intelligence. Menlo Park,

Calif.: American Association for Artificial Intelligence.

Ndumu, D.; Nwana, H.; Lee, L.; and Haynes, H. 1999. Visualization and Debugging of Distributed Multiagent Systems. *Applied Artificial Intelligence Journal* 13(1): 187-208.

Quinlan, J. 1994. *C4.5: Programs for Machine Learning*. San Francisco, Calif.: Morgan Kaufmann.

Tambe, M. 1997. Toward Flexible Teamwork. *Journal of Artificial Intelligence Research* 7:83-124.



Milind Tambe is a project leader at the Information Sciences Institute, University of Southern California (USC), and a research associate professor with the Computer Science Department at USC. He received his Ph.D. in 1991 from the School of Computer Science at Carnegie Mellon University. His

interests are in the areas of multiagent systems, specifically multiagent teamwork, negotiation, agent modeling, and real-time performance, and he has published extensively in these areas. He is currently on the editorial board of the *Journal of Artificial Intelligence Research*, the *Journal of Autonomous Agents and Multiagent Systems*, and *IEEE Intelligent Systems*. He is the program cochair of the International Conference on Multiagent Systems and has served as finance and local arrangements chair for the International Conference on Autonomous Agents (Agents) and as the senior program committee member for the National Conference on Artificial Intelligence as well as for Agents. He is also a member of the board of trustees for RoboCup, the Robot World Cup Federation.

Stacy Marsella is a researcher at the University of Southern California Information Sciences Institute and is a member of the Center for Advanced Research in Technology for Education. His research interests include virtual humans, agent-based interactive drama, agent teamwork, and the analysis of both human and agent teamwork. He received his Ph.D. from Rutgers University.

Case-Based Reasoning

Experiences, Lessons, & Future Directions

Edited by David B. Leake



Case-based reasoning is a flourishing paradigm for reasoning and learning in artificial intelligence, with major research efforts and burgeoning applications extending the frontiers of the field. This book provides an introduction for students as well as an up-to-date overview for experienced researchers and practitioners. It examines the field in a “case-based” way, through concrete examples of how key issues — including indexing and retrieval, case adaptation, evaluation, and application of CBR methods — are being addressed in the context of a range of tasks and domains. Complementing these case studies are commentaries by leading researchers on the lessons learned from experiences with CBR and visions for the roles in which case-based reasoning can have the greatest impact.

ISBN 0-262-62110-X

420 pp., index.

The AAAI Press • Distributed by The MIT Press

Massachusetts Institute of Technology, Cambridge, Massachusetts 02142

To order, call toll free: (800) 356-0343 or (617) 625-8569.

MasterCard and VISA accepted. <http://www.aaai.org>