Agent Assistants for Team Analysis

Taylor Raines, Milind Tambe, Stacy Marsella
Information Sciences Institute and Computer Science Department
University of Southern California
4676 Admiralty Way, Suite 1001
Marina del Rey, CA 90292, USA
{Raines, Tambe, Marsella}@isi.edu

Abstract

With the growing importance of multi-agent teamwork, tools that can help humans analyze, evaluate, and understand team behaviors are becoming increasingly important as well. To this end, we are creating ISAAC, a team analyst agent for post-hoc, off-line 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 via machine learning over teams' external behavior traces, where the specific learning techniques are tailored to the particular model learned. Additionally, ISAAC employs multiple presentation techniques that can aid human understanding of the analyses. This paper presents ISAAC's general conceptual framework and its application in the RoboCup soccer domain, where ISAAC was awarded the RoboCup scientific challenge award.

1. INTRODUCTION

Multi-agent teamwork is an important area of agent research, with a growing number of applications, including multi-robotic 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 multi-agent 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 been largely 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 multi-agent visualization systems, such as [Ndumu et al 1999], has been motivated by multi-agent 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 multi-agent team behaviors by (i) locating key aspects of team behaviors that are critical in team success or failures; (ii) diagnosing such team behaviors, particularly, problematic behaviors; (iii) suggesting alternative courses of action; and (iv) 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, while 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, and 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: an individual agent action, agent interactions, and overall team 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 that action that played a prominent role in its success or failure, but a similar presentation would be incongruous for a global analysis, since no single action would suffice. Global analysis requires a more comprehensive explanation that ties together seemingly unconnected aspects and trends of team behavior. ISAAC uses a natural language summary to explain the team's overall performance, using its multimedia viewer to show examples where appropriate. The content for the summary is chosen based on ISAAC's analysis of key factors determining
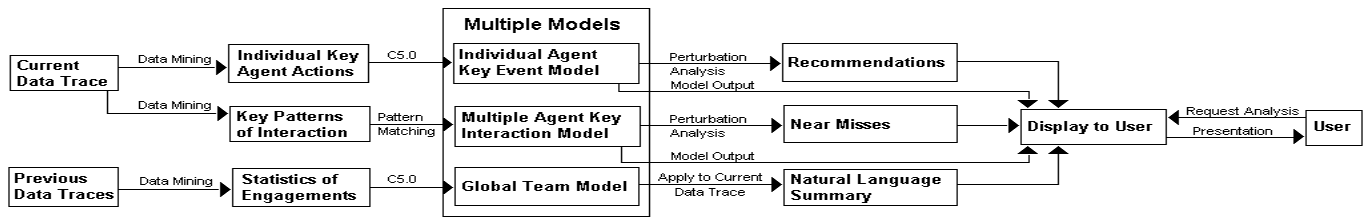
Figure 1: Flow Chart for ISAAC Model Generation and Analysis

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 multi-agent research. Here, two teams of 11 agents each play soccer in a noisy, dynamic environment. Agent-team analysis is critical in RoboCup since team developers wish to understand the strengths and weaknesses of teams and to improve such teams. Indeed, ISAAC has been applied to all of 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. ISAAC is available on the web at http://coach.isi.edu.

2. ANALYSIS USING MULTIPLE MODELS

ISAAC's analysis can be split in to two phases of *model acquisition* and *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 extension to C4.5, to create rules of success or failure [Quinlan 1994]. For analysis of agent interactions (*multiple agent key interaction model*), pre-defined 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.

Utilizing the models involves using 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 (See figure 2). A *perturbation analysis* is then performed to recommend changes to the team by changing the rule condition 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.

Figure 2: ISAAC's 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. Below is 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).

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 very hard against the HAARLEM defense, <u>keeping the ball in their half of the field for 84% of the game</u> and allowing ample scoring opportunities. HAARLEM <u>pulled their defenders back to stop the onslaught</u>, but to no avail. To that effect, 11monkeys <u>was able to get past HAARLEM's last defender</u>, creating 2 situations where only the goalie was left to defend the net. 11monkeys also handled the ball better, <u>keeping control of the ball for 86% of the game</u>. HAARLEM <u>had a tendency to keep the ball towards the center of the field</u> as well, which may have helped lead them to ruin given the ferocity of the 11monkeys attack.

3. 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 instance, ISAAC found a problematic pattern in the shooting behavior of Andhill'97, *the 2nd place winner of '97*. Not only was this surprising to us, but this 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 Andhill team in the future." [Andou, personal communication]

Another interesting result from the individual 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 4 teams (from 9 rules to 5 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 due to the fact that 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 that 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, e.g., some early games were unrepresentative and some teams were changed by hand during the competition. Yet in practice, ISAAC's predictive accuracy was 70% 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 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% 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% of the participants rated the summaries as *very good*, with another 10% 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% of the games were classified correctly, resulting in an error rate of 13%. Our test set of (unseen) RoboCup '99 games produced 72% classified correctly, for an error rate of 28%. 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 may 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 the 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.

4. CONCLUSION

With the growing importance of multi-agent teamwork, it is now increasingly critical to build automated assistants to aid developers in analyzing agent team behaviors. We have taken a step towards this goal, by building an agent called ISAAC for post-hoc, off-line agent-team analysis. ISAAC uses two key novel ideas in its analysis. First, ISAAC uses multiple models of team behavior to analyze different granularities of agent actions, using inductive learning techniques, enabling the analysis of differing aspects of team behaviors. 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. While ISAAC has currently been applied in the context of the RoboCup soccer simulation, we hope to apply it to other domains, such as battlefield simulations [Tambe 97] in the future.

REFERENCES

[Johnson 1994] Johnson, W. L. Agents that Learn to Explain Themselves. In *Proceedings of AAAI-94*, 1994.

[Ndumu et al. 1999] Ndumu, D., Nwana, H., Lee, L., Haynes, H. Visualization and debugging of distributed multi-agent systems. In *Applied Artificial Intelligence Journal, Vol 13 (1),* 1999.

[Quinlan 1994] Quinlan, J. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1994.

[Tambe 1997] Tambe, M. Towards Flexible Teamwork. In *Journal of Artificial Intelligence Research, Vol. 7,* 1997.