

Trying to Understand RoboCup

Kumiko Tanaka-Ishii, Ian Frank, and Katsuto Arai

■ As the English striker Gary Lineker famously said, “Football is a very simple game. For 90 minutes, 22 men go running after the ball, and at the end, the Germans win.” Although the game is simple, analyzing it can be hard. Just what makes one team better than another? How much difference do tactics make? Is there really such a thing as a “lucky win”? Here, we try to answer these questions in the context of RoboCup. We take the giant set of log data produced by the simulator tournaments from 1997 to 1999 and feed it to a data-munching program that produces statistics on important game features. Using these statistics, we identify precisely what has improved in RoboCup and what still requires further work. Plus, because the data muncher can work in real time, we can also release it as a proxy server for RoboCup. This proxy server gives all RoboCup developers instant access to statistics while a game is in progress and is a promising step toward an important goal: understanding RoboCup.

Before the real soccer world cup in France in 1998, Simon Barnes (1998) wrote in the *Times*:

“The basic error in trying to understand the World Cup is to believe that it is all about trying to find the best football team in the world. It is not.... Football teams have their being in four dimensions, and the fourth dimension is Time. What we seek is the best football team in the world between June 10 and July 12.”

Just as it is for the real World Cup, so it is with RoboCup. The annual tournaments are exciting and unpredictable. They produce winners and losers. They produce drama.

However, it would be wrong to believe that the winning teams in RoboCup are always the “best.” There are too many imponderables. The fortunes of a team can turn on an interpretation of the rules, problems with camera calibrations, difficulties with computer networks, or

accidental damage to components—not to mention plain, old-fashioned luck.

In real sport, unpredictability brings with it excitement. The underdogs can always turn the tables, and the losers can always cry “we was robbed.” However, RoboCup is science. How should the results of the games be interpreted?

We were faced with this question in the earliest RoboCups, when we began work on an automated commentator for the software league. Human commentators habitually comment on “slices of luck” and “the run of play.” To mimic such commentary using a computer, we realized we needed an “expert analysis module” that could analyze how well teams were actually playing. The commentator system we built around this module was called MIKE (multiagent interactions knowledgeably explained).

Early on, we realized that expert analysis should have other uses within RoboCup. We wrote about these uses for the 1998 RoboCup workshop, where we included the picture in figure 1.

What we envisaged was a module that could take the low-level events reported by the simulation league’s SOCCER SERVER (Noda et al. 1998) and produce an analysis of each team’s playing style, tactics, strengths, and weaknesses. As well as using this module in MIKE, we hoped other researchers would find it useful for tackling the specific challenges of learning, teamwork, and opponent modeling laid out in Kitano et al. (1997).

As a first step toward a genuine expert analysis module, we decided to focus on statistics. In the year between the 1998 and the 1999 RoboCups, we concentrated our efforts on improving the range of statistics compiled by MIKE (so much so that people listening to a MIKE commentary sometimes complained that it sounded like an American baseball commentator!).

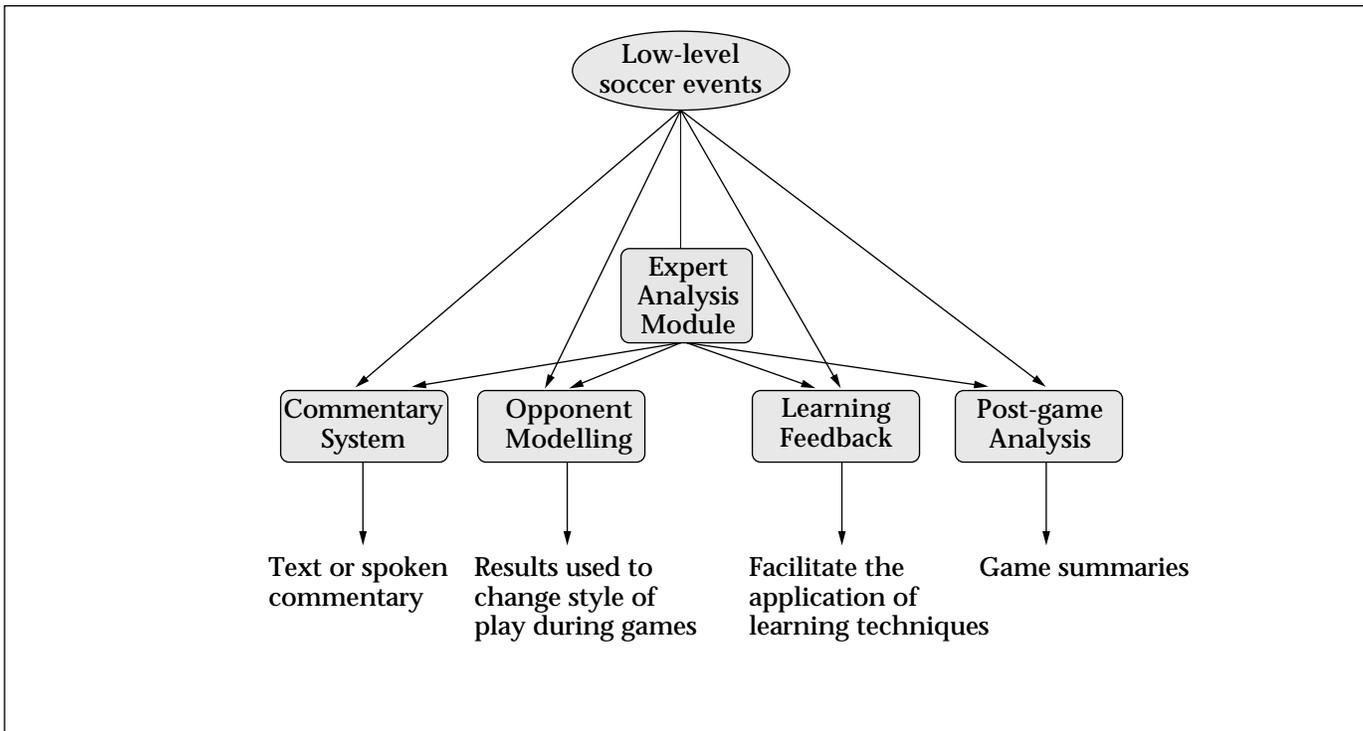


Figure 1. Likely Uses of an Expert Analysis Module.

A statistics proxy server based on MIKE's analysis modules should be released very soon (certainly by the time this issue of *AI Magazine* is in print). Although the scope of MIKE's analyses does not yet compare with commercial programs developed for human coaches and teams (for example, see SoftSport),¹ there are some novel ideas. For example, the first version of the server will include the representation of ball-play chains as first-order Markov processes and the calculation of players' defensive areas with Voronoi diagrams (Tanaka-Ishii et al. 1998)

We hope the STATISTICS SERVER will be well used in 2000 and beyond, eventually developing into a full-fledged expert analysis module. Real-time high-level analysis of SOCCER SERVER games should really come into its own with the introduction of a touchline coach client (planned for 2000). This client will make it possible to use statistics for online learning during a competitive game and hugely expand the scope for applying opponent modeling (for example, identifying the opponent's key players, passwork patterns, and styles of play) and improving teamwork (for example, identifying players not fulfilling their assigned roles, enabling appropriate adaptations to be made).

Already, though, we have used MIKE's statistics to demonstrate the improvement in RoboCup skills to date. It is simple to observe

that the best simulator teams from any year's RoboCup beat the best teams from the year before. We wanted to know why this result was true. For the 1999 RoboCup workshop, we wrote a paper that used MIKE's statistics to analyze the first two RoboCups. In the following discussion, we reprise some of this analysis, taking the chance to update it to include the results from the 1999 tournament.

Analyzing the RoboCups, 1997 to 1999

The SOCCER SERVER generates a lot of data. The size of the data files for the 243 complete games in the first 3 RoboCups is over 500 megabytes. We used MIKE's statistics to analyze these logs, focusing on the differences between the winning and losing teams.

Team designers are allowed to manually change their team settings and code at half-time, so we separately collected the statistics for the winners and losers of each half-game (excluding the 97 half-games that were drawn).

We analyzed the log files in terms of a set of 32 features (Tanaka-Ishii et al. 1999) that describe different aspects of a soccer game. Here, we pick out just a few features and give graphs that show how they have changed over the three RoboCups. For reasons of space, we

focus here primarily on the graphs for winning teams. In each of these graphs, the x axis is partitioned into 30 equal and contiguous intervals that span the observed maximum and minimum values of the feature in question. To plot a line for the results of any given year, the percentage of games for which the value of the feature falls within each interval is counted, and a smoothed curve is fitted through the 30 points.

Formation

In real soccer, the entire formation of a team follows the movement of the ball across the field. To measure this effect in RoboCup, we defined the compactness of a team as the X distance between its front-most player and its rear-most player (excluding the goalkeeper). Teams with lower compactness tend to be more dynamic and benefit from having more players closer to the ball. Figure 2 shows the distribution of the average compactness of winning teams, demonstrating clearly that compactness has improved each year (the plots move to the left). This improvement is partly the result of the introduction of the off-side rule in 1998. *Off-sides* allow a team to protect itself against counterattack by moving upfield as a unit with the ball. As more teams take advantage of this rule, it becomes an aspect of strategy that developers cannot afford to neglect.

However, figure 3 shows that the variance of compactness during games has remained largely unchanged over the three years. Thus, although team formation has improved to dynamically follow the ball, the degree of compactness does not yet change significantly during a game.

Player Correlations

Another way to analyze players' movements is in terms of correlations between player locations. We wrote code to calculate these correlations in the form of two 22×22 symmetric matrices: one for the correlation of players' X locations and the other for the players' Y locations. These matrices are generated for each time step of the game. Figures 4 and 5 shows the average of the values in the X and Y correlation matrices over entire half-games. These averages are calculated for the entries in the 11×11 submatrices giving correlations between teammates (there were no negative elements in these submatrices, making the average a realistic measure). Although the Y correlations show no significant change, the X correlations show a substantial improvement, especially in the movements of the 1999 winners.

Other things that can be demonstrated with correlations are the extent to which a team's

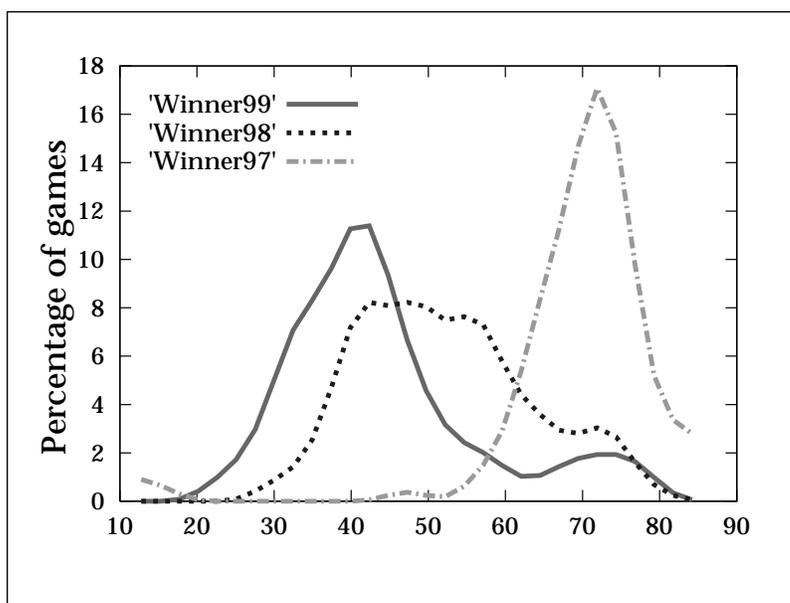


Figure 2. The Average of Compactness.

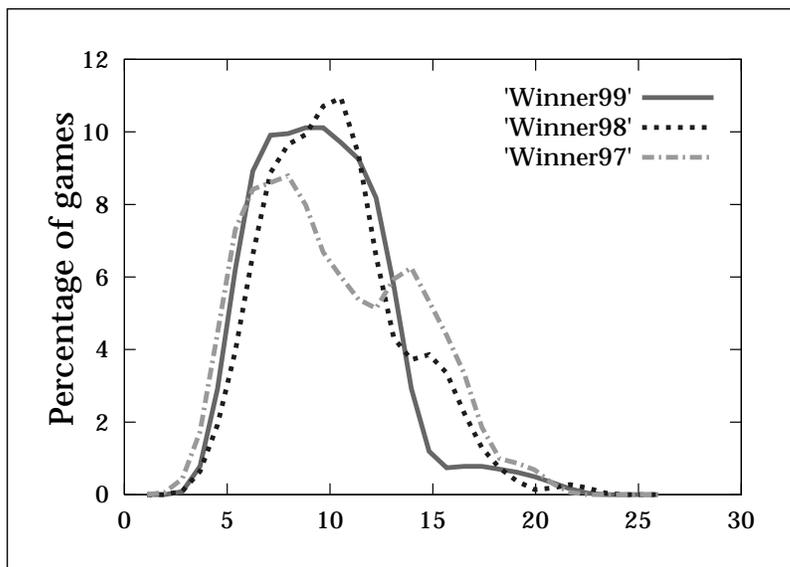


Figure 3. The Variance of Compactness.

defenders or attackers function as a unit and the amount of marking in a game. For example, figure 6 shows the X correlation values for a team's front three defenders, and figure 7 shows the average of teams' X correlations with their opponents. Again, these results show a marked improvement in the winning teams of 1998 and 1999 (these graphs are further to the right). Note that in the case of marking, the correlation matrices between winners and losers are symmetric, so we cannot say which team's players are actually doing the marking. In general, this problem is not trivial—we are investigating how to collect more statistics on

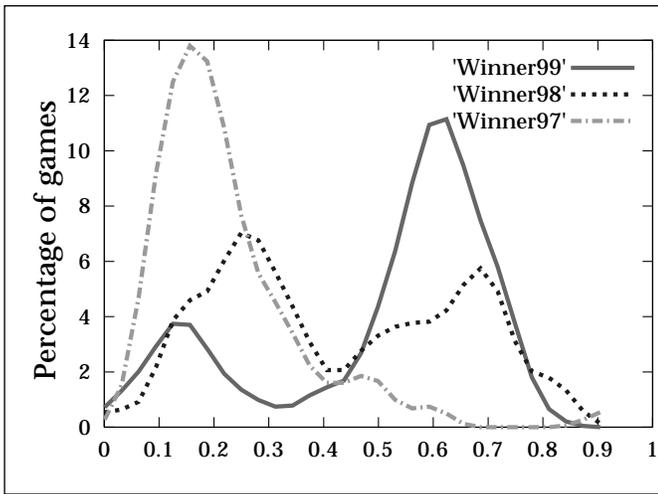


Figure 4 (left). The Average X Correlations.

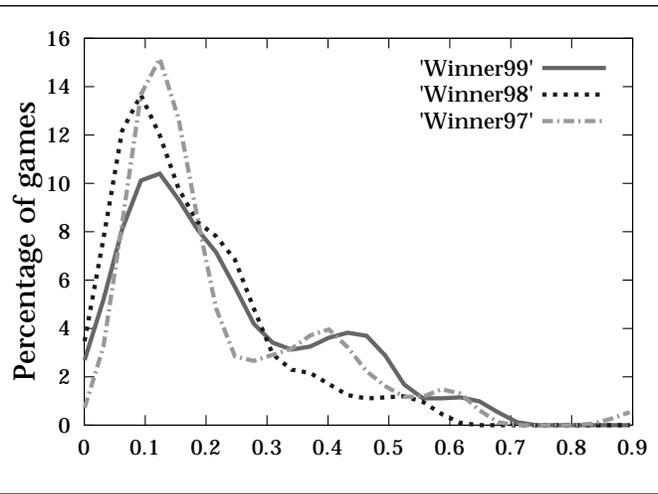


Figure 5 (right). The Average Y Correlations.

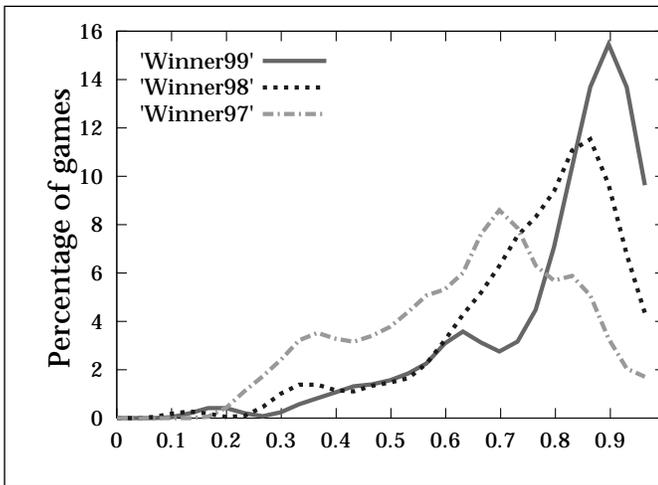


Figure 6 (left). The Average X Correlation among Attackers.

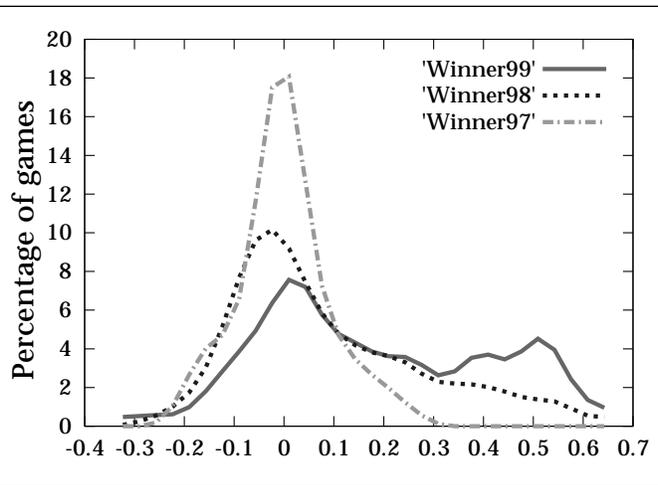


Figure 7 (right). The Average X Correlation between Winners and Losers (Marking).

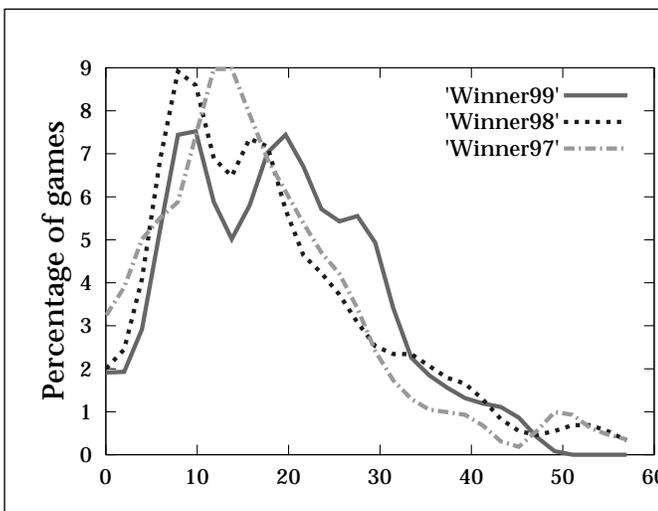


Figure 8 (left). The Total Number of Passes.

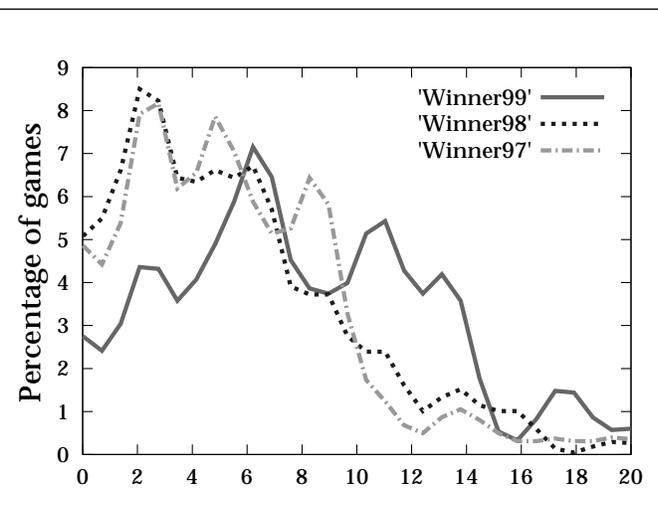


Figure 9 (right). The Total Number of Dribbles.

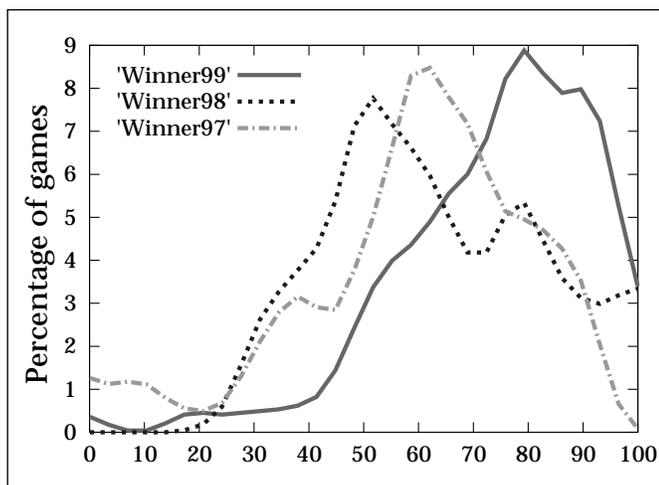


Figure 10 (left). Average Pass Success Rate of Winners.

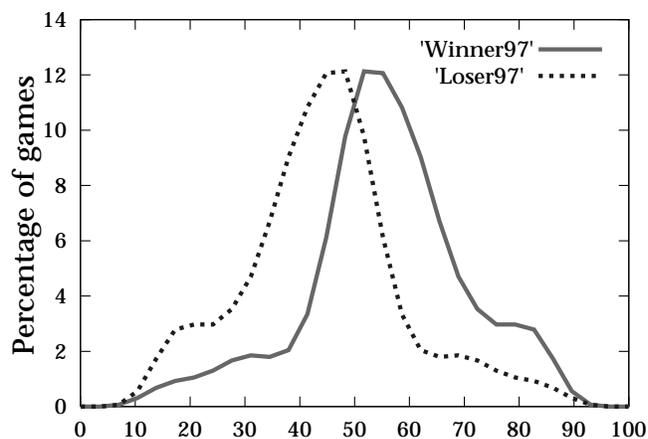


Figure 11 (right). Possession Rates of Winners and Losers in 1997.

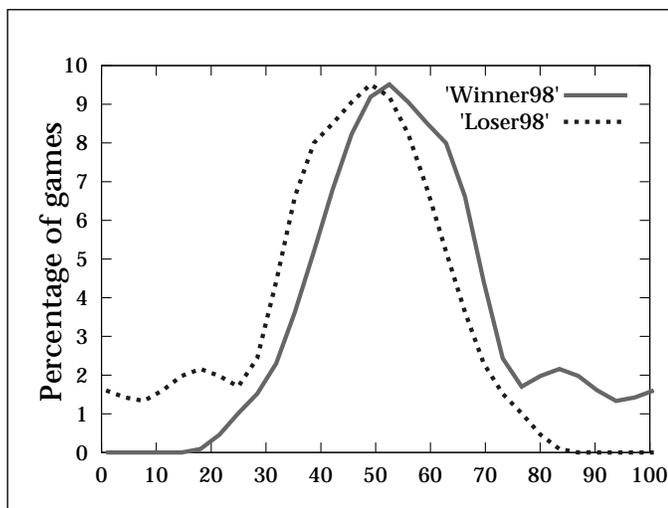


Figure 12 (left). Possession Rates of Winners and Losers in 1998.

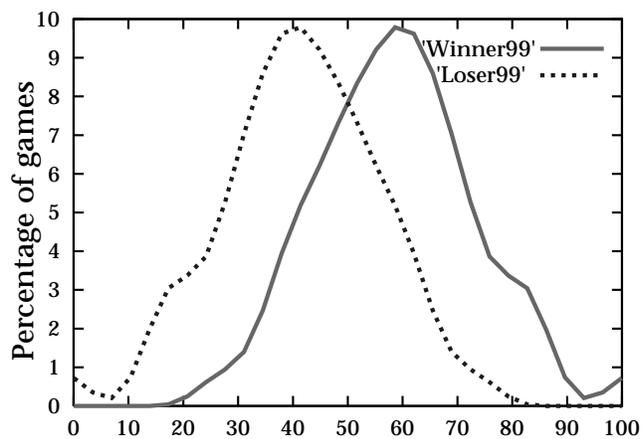


Figure 13 (right). Possession Rates of Winners and Losers in 1999.

marking by examining the dependencies between the directions and the timings of players' movements.

A general trend with all the correlation-based statistics we examined was that the Y correlations showed no noticeable improvement over the three RoboCups. Interestingly, though, we found that the 1997 figures were skewed by a small number of games that exhibited what we called the *rugby effect*: teams simply surrounding the ball with their players in a scrum that slowly moved across the field. Motivated by the introduction of the off-side rule, teams have improved their X correlations to the extent that the 1999 average is higher than the biased 1997 value. For the Y correlations, we expect such improvements to come with further research.

Ball-Handling Skills

Drawing graphs of statistics didn't reveal much about changes in basic ball-handling skills over the first two RoboCups. Thus, for RoboCup-99, we utilized *principal component analysis* to show that an underlying ability to handle the ball well was important for success. The graphs for the most recent RoboCup, however, show slightly clearer trends. For example, in figures 8 and 9, the 1999 winner's plots of total passes and dribbles are slightly further to the right. More compellingly, figure 10 shows that winning teams in 1999 had much better pass success rates.

Finally, figures 11 to 13 show that in 1999, the gap between the possession rates of the winners and the losers of games started to increase, indicating that in 1999, there was

more of a link between ball-handling skill and performance; however, in previous years, teams still had reasonable chances of winning even when they lacked basic skills.

Conclusions

We picked out some of the most compelling statistical evidence demonstrating the soccer improvements over three years of the RoboCup competition. A brief summary of some of the main points might list the following five elements: First, teams now work together to follow the play better (despite having to manage stamina more carefully because of changes in the server rules each year). Second, teamwork has improved significantly along the long side of the field but not along the short. Third, improved teamwork can be seen in the way that defenders (and forwards) share their roles. Fourth, teams now carry out more marking. Fifth, the link between ball-handling skills and performance has strengthened.

Can statistics help to understand RoboCup? Yes, in many ways. Thorough analysis can highlight the aspects of play that have improved, demonstrating actual progress. At the same time, by highlighting what has not improved, statistical analysis can enable researchers to identify the most promising directions for further work. In addition, a statistics server can be used to inform a commentary system, model opponents, carry out learning, and do post-game analysis.

As RoboCup progresses, we expect statistical analysis to become even more significant. Most obviously, teams will not be able to ignore the advantage that comes with effective, real-time use of the online coach. We are also working on adapting our code to work with servers that produce logs from video footage of the real robot leagues.

Further, we can predict that the skill levels of teams will start to converge as a result of code being released into the public domain and the body of RoboCup research becoming larger. When this research expands, statistics will be indispensable for distinguishing between closely matched teams.

As a RoboCup developer, if I develop a new approach for technique *X*, how do I test it? Ideally, I'd like a black box that can test a team with technique *X* against a team without and give me a 95-percent assurance that *X* either weakens or strengthens different aspects of the team's play. For RoboCup-99, we suggested that the ultimate use of statistics in this way would be to develop a team based on a common model that could pool the efforts of many researchers by allowing them to submit different modules.

For all these reasons, we believe that our statistics proxy server is just a first step. We hope it will form the core for a significant expert analysis module incorporating techniques developed by many RoboCup researchers. A good starting point would be the work done by the developers of ISAAC, described elsewhere in this issue.

Note

1. SECOND LOOK soccer analysis software by SoftSport Inc. can be found at www.soft-sport.com/.

References

- Barnes, S. 1998. If You're Hot You're Hot, If You're Not You're Out. *The Times*, London, 8 June.
- Kitano, H.; Tambe, M.; Stone, P.; Veloso, M.; Coradeschi, S.; Osawa, E.; Matsubara, H.; Noda, I.; and Asada, M. 1997. The RoboCup Synthetic Agent Challenge 97. In Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence (IJCAI-97), 24-29. Menlo Park, Calif. International Joint Conferences on Artificial Intelligence.
- Noda, I.; Matsubara, H.; Hiraki, K.; and Frank, I. 1998. SOCCER SERVER: A Tool for Research on Multiagent Systems. *Applied Artificial Intelligence* 12(2-3): 233-251.
- Tanaka-Ishii, K.; Frank, I.; Noda, I.; and Matsubara, H. 2000. A Statistical Perspective on the RoboCup Simulator League: Progress and Prospects. In Proceedings of the Third International Workshop on RoboCup. Berlin: Springer-Verlag. Forthcoming.
- Tanaka-Ishii, K.; Noda, I.; Frank, I.; Nakashima, H.; Hasida, K.; and Matsubara, H. 1998. MIKE: An Automatic Commentary System for Soccer. In Proceedings of the International Conference on Multiagent Systems, 285-292. Washington, D.C.: IEEE Computer Society Press.



Kumiko Tanaka-Ishii is an assistant professor, Faculty of Engineering, University of Tokyo, Japan. Her research interest is the dynamic knowledge explanation of evolving contexts, which concerns domains

such as real-time natural language interface and dynamic knowledge extraction. Her recent research projects involve automatic commentary, navigation, and presentation systems. She received her Ph.D. (cross-lingual ambiguity resolution in natural language processing, 1997), her M.E. (concurrent functional programming, 1993), and her B.E. (applied statistics, 1991) from the University of Tokyo Japan. From 1995 to 1996, she was an invited researcher at the LIMSI-CNRS in France. She worked for Electrotechnical Laboratory, Japan, from 1997 to 2000. Her e-mail address is kumiko@ipl.t.u-tokyo.ac.jp.



Ian Frank is a European Union Science and Technology Research Fellow and is currently based at the Complex Games Lab at the Electrotechnical Laboratory, Tsukuba, Japan. His research interests include game theory, search, planning, and automated explanation in complex domains. He holds a Ph.D. in AI from Edinburgh University (1996, U.K. Distinguished Dissertation Award); an M.Sc. in knowledge-based systems, also from Edinburgh University (1991); and a B.Sc. in mathematics from Durham University (1988). His e-mail address is ian@etl.go.jp.

search, planning, and automated explanation in complex domains. He holds a Ph.D. in AI from Edinburgh University (1996, U.K. Distinguished Dissertation Award); an M.Sc. in knowledge-based systems, also from Edinburgh University (1991); and a B.Sc. in mathematics from Durham University (1988). His e-mail address is ian@etl.go.jp.



Katsuto Arai is a fourth-year student at the University of Chiba. His interests include network hardware and software design. His e-mail address is karai@cogsci.l.chiba-u.ac.jp.