USING A DATAWAREHOUSE TO EXTRACT KNOWLEDGE FROM ROBOCUP TEAMS

Isabel Gonzalez, Pedro Abreu and Luis Paulo Reis
FEUP - Faculdade de Engenharia da Universidade do Porto
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
DEI - Departamento de Engenharia Informatica, Rua Dr. Roberto Frias
4200-465 Porto, Portugal
LIACC - Laboratório de Inteligência Artificial e Ciência de Computadores
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
aisbel@fe.up.pt, pha@fe.up.pt, lpreis@fe.up.pt

Keywords: Knowledge Extraction, DataWarehouse, Game Analysis, Robotic Soccer, RoboCup, Simulation.

Abstract: RoboCup is a scientific and educational, international project that involves artificial intelligence, robotics and sport sciences. In these competitions, teams of all around the world participated in distinct leagues. In the begining of the Coach competition, one of RocoCup leagues,the goal of the researchers was to develop an agent (Coach that provides advices to teammates about how to act and with improve team performance. Using the resulting improved coach agent, with enhanced statistic calculation abilities, a huge amount of statistical data was gathered from the games held at Bremen 2006. This data was then stored and treated in a data warehouse system obtaining a good high level perspective/knowledge of the RoboCup simulated soccer tournament. According to the results, the team that represented our country, has a much more goal opportunities in comparison with the majority of the teams, but this team did not score many goals. In terms of more occupied regions, the best four teams in the tournament did not occupy many times the left and right wings, compared to others regions. In the future the our country team needs to develop new strategies that use these two areas preferrentialy in order to achieve better results.
.

1 INTRODUCTION

RoboCup is a scientific and educational, international project that involves artificial intelligence, robotics and sport sciences. The main goal of RoboCup is to develop a team of fully autonomous humanoid robots that can defeat the human world champion team in soccer by 2050.

Since the first official RoboCup games in 1997, in which over 40 teams participated (real and simulation combined), there have been lots of other international and regional competitions and scientific conferences, related with it. In these competitions, teams of all around the world participated in several distinct leagues and researchers presented scientific work about their approaches to several sub problems.

Nowadays there are many different leagues divided into four main classes: RoboCup Soccer, RoboCup Rescue, RoboCup@Home and RoboCupJunior. In these classes, there are also several sub leagues. For example in RoboCup Soccer competitions there are many sub leagues like Soccer Simulation leagues (2D, 3D and Coach) and robotic leagues (small-size, middle-size legged and humanoid). One of the competitions present in the Soccer simulation leagues is the Coach competition in which research groups have to develop a Coach Agent capable of high-level action and intelligent analysis of the simulated soccer game. In the beginning of this competition, the goal of the researchers in this league, was to develop an agent (Coach) that provides advices to teammates about how to act and with this improve team performance. The communication system between the coach agent and his players was supported by a standard coach language called CLang (Federation, 2007). This language was developed based on COACH UNILANG (Reis and Lau, 2002) language created by FC Portugal in 2001 which enables high-level communication between a coach agent (or human coach) and a soccer robotics playing team. Clang started to be a very simple language but evolved into a very complete

language for coaching soccer teams, containing most of the features introduced originally in COACH UNILANG.

The main goal of the coach competition was changed after its first four editions (2001-2004) transforming the competition into a game analysis challenge (Kuhlmann et al., 2005). The league has been used, since then, to extract usefull information about the game. In the context of this league, in the past years, many advances have been made in these research topics: Development of team opponent models (players individual models and teams collective models (Kuhlmann et al., 2006); Methodologies for using Coaching to help teams to improve their performance in the simulated robotic soccer domain (Riley et al., 2002); Adjustable autonomy in real-world multi-agent environments (Scerri et al., 2001).

The main goal of this research is to extract the highest knowledge about the way of playing of the teams that participated in the 2D simulation league of RoboCup World Championship 2006 held in Bremen, Germany. Using the available coach program of FC Portugal team (team, 2007) as a base, we have implemented new algorithms for calculating several statistics. Using the resulting improved coach agent, with enhanced statistic calculation abilities, a huge amount of statistical data was gathered from the games held at Bremen 2006. This data was then stored and treated in a data warehouse system obtaining a very good high-level perspective/knowledge of the RoboCup simulated soccer tournament.

2 PROJECT ARCHITECTURE

In order to achieve the best knowledge of the playing style of all robotic soccer teams participating in the last simulation league tournament of RoboCup, this project used the architecture illustrated in figure 1.

In RoboCup 2006, as in others similar tournaments, many teams compete together to achieve the best result having as main goal, obviously, to win the competition. In order to start the simulation the players (agents) need to connect to a server (soccerserver) and after that the simulation could be started. At the final of the game, the soccer server generates the corresponding log file, which characterized the entire game, including the players and ball positions and velocities and other relevant information. The first phase (1) of this project consisted in collecting all log files from RoboCup 2006 tournament. In the next two phases (2 and 3), the coach agent (introduced in the previous section), using the log files then extracts the maximum information about the game and calculates different types of statistics e.g. number of successful passes, number of goals etc. After that, in phase 4, the Data Warehouse (DW) was created and the statistics, that were previously calculated, were transformed in order to be easier to store in the DW. The statistics were then stored in an appropriate format in the Data Warehouse. In the final project phase (5), after the storage of all tournament data in the DW, using SQL queries, we aimed at extracting all kinds of useful high-level information about all robotic teams individual and collective playing style.

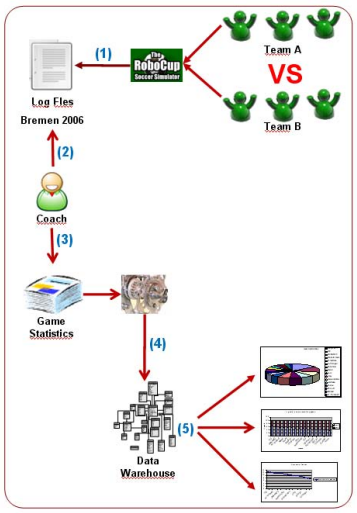

Figure 1: Project's Architecture.

3 GAME STATISTICS

According to the Coach Competition rules, the coach has to analyze log files of a given team and discover its playing pattern. The main challenge is to use the low-level information contained in the log files (positions and velocities) to calculate and appropriately use high-level information of the match. The games statistics calculated included: passes (and its results), ball losses and recoveries, game situation occurrences (corners, free-kicks, among others), goals and scoring opportunities.

3.1 Field Regions

With the purpose of extracting information about the places where there is a higher occupation rate for both teams in a game, we have separated the field in eight regions: our middle, their middle, our penalty box, their penalty box, our left wing, their left wing, right wing and their right wing.

The region delimiters are the following 2 for left team, and the opposite of right team.

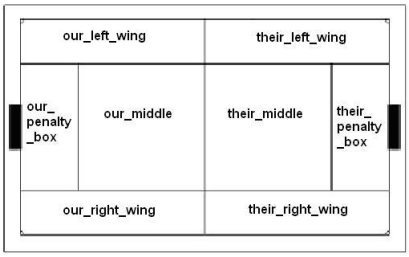

Figure 2: Considered Field Regions.

These regions are built giving the points of a rectangle that defines each region in the field. The results depend on the team analyzed, the side that team starts to play and if you are in the first half or the second half of the game. All these statistics are saved in a final game file that will be then analyzed.

3.2 Pass Information

A pass may be considered when a player kicks the ball with enough force in a specific direction, aiming to a teammate that is supposed to receive the ball.

Regarding passes, the statistics calculated are the total of passes, the percentage of passes in each part of the game, the percentage of passes of each team, the same with each player (passes executed and received), the percentage of successful passes in each part of the game of each team. Pass statistical information is also calculated by each field region.

3.3 Ball Recoveries and Losses

A ball recovery occurs when a player from the team has the ball possession in a given cycle (e.g. may kick the ball) and he losses the ball possession for a member of the opponent team (a player from the opponent team may kick the ball). This statistic is the opposite of a ball recovery. When a team is able to do a ball recovery, the other team looses the ball possession.

Regarding this statistic, the system calculates the total of ball recoveries and losses, the percentage of recoveries and losses in each part of the game, the percentage of recoveries and losses of each team, the same with each part of the game. Statistics are also calculated by each field region.

3.4 Game Occurrences

This game parameter summarizes the occurrences that happened in the game. These occurrences can be free kicks (right or left), goals (right or left), offsides, corners, etc. In this situation we calculate the total occurrences, the percentages of each occurrence and the percentage of occurrences in each region.

3.5 Goal Scoring and Opportunities

A goal scoring opportunity occurs when an attacking player has the ball possession in a dangerous area. Dangerous area is a subjective concept that we estimate as a region sufficiently near the opponents goal that a shoot may be successful. In this statistic calculation we estimated the total of goal opportunities, the percentage of opportunities in each part of the game, the percentage of opportunities of each team and player and the same with the halfs of the game. Besides this, it is also calculated the percentage that each region is visited when there is an opportunity.

4 DATA WAREHOUSE

In order to get information about statistics, the final game file was split into several different files, one for each type of statistic. In each file the data was separated by tabs considering the future storage a data warehouse. Although it could be a good choice, XML is not being considered as a storage format because it implies defining a schema and then validating each file with this schema and, in the process of storage, its use would decrease the performance of all process comparing with the tabs separation method.

The reason for using a data warehouse instead of a conventional database is mainly because, in this project, the goal was to collect a higher number of files at once and after that do some interrogations in the data warehouse. The performance of the data warehouse in terms of time delay is also critical so the data warehouse seems the best option in this case. In order to represent most of the interesting soccer related statistics the following data warehouse was been defined.

In this structure three distinct dimensions were defined: Participation, Game Summary and Classification. The Participation is the dimension that had all information about the participants of the game like referee, trainer and player, filtered by data and league. The Game Summary dimension is referred to all game statistics calculated per team like the number of attacks, number of assists, ball recovery among others. Finally, the Classification dimension has information about the final results that teams achieve in a specific competition. In this project only the game summary dimension have been used mainly to simplify

the analysis that would be confusing using all kinds of data for all dimensions, for all the teams.

5 RESULTS AND DISCUSSION

For a better understanding this section is divided in three parts.

5.1 Goal Opportunities

Doing a comparison between the number of goals scored and the percentage of goal opportunities the Ri-one team was the team with the best performance. This team achieved more than thirty-five goals with only four percent of the total goal opportunities. In the opposite side teams like FC Portugal needed seven percent of goal opportunities to scored only four goals. Analyzing the games it is easy to see that Ri-One team had a simple attacking behavior, based on a long pass to the team forwards that resulted in a simple goal against most of the teams.

5.2 Successful Passes

This statistic of the game must be improved for all the teams in the near future. The values in it were between eight to thirteen percent which is a very low value compared, for example, with real football. Teams use a very fast but not so safe playing style, resulting in lots of ball losses when executing passes.

5.3 Regions Visited by the Best Teams in the Tournament

The area of the field more occupied by these teams were the two midfields area (our and their). Peculiar is the behavior of the Ri-one team, which occupied the penalties boxes areas rarely (compares with other field area). Other interesting observation was the behavior of the Wright Eagle (WE) team, which occupied the opponent left wing area less than one percentage of their total

6 CONCLUSIONS AND FUTURE WORK

This paper describes an approach to the use of a data warehouse to extract knowledge from RoboCup teams playing style. The transformations and data extraction were developed in a very easy way by using the data warehouse, enabling to use the results achieved to get information about the teams and their playing style. With these results we concluded that teams like TokyoTech or FC Portugal needed many opportunities to score a goal in contrast with Ri-one that scored most of their opportunities. An interesting observation is that FC Portugal team has much more goal opportunities in comparison with other teams, but this team did not score many goals. So, this seems like a parameter to improve in a near future by the FC Portugal team, for example, shooting more to the goal or developing simple but effective moves in the opponents area, to score goals.

The next developments in this project shall focus two distinct areas. In the future the data warehouse will have data not only from RoboCup 2006 Competition but also from previous tournaments. This point will improve the knowledge about all the participating teams, which will also enable to raise the performance of a given Team in a near future. This point could also allow to store a higher amount of data throw the use of the developed data warehouse.

The other area of work, regards the implementation of new types of statistics in order to better characterize the game.

REFERENCES

Federation, R. (2007). Clang: The robocup coach language. In *Avaiable online at http://www.cs.utexas.edu/ ywwong/wasp/robocup-clang.html*.

Kuhlmann, G., Knox, W., and Stone, P. (2006). Know thine enemy: A champion robocup coach agent. pages 1463–1468.

Kuhlmann, G., Stone, P., and Lallinger, J. (2005). The ut austin villa 2003 champion simulator coach: A machine learning approach. pages 636–644. Springer Berlin / Heidelberg. ISBN 978-3-540-25046-3.

Reis, L. and Lau, N. (2002). Coach unilang a standard language for coaching a (robo) soccer team. Lecture Notes in Computer Science, Volume 2377, pages 183–192. Springer-Verlag, Berlin. ISBN 3-540-43912-9.

Riley, P., Veloso, M., and Kaminka, G. (2002). An empirical study of coaching. pages 215–224. Springer-Verlag.

Scerri, P., Pynadath, D., and Tambe, M. (2001). Adjustable autonomy in real-world multi-agent environments. in agents. pages 107–116. ISBN 978-1-57735-131-3.

team, F. P. (2007). Fc portugal team oficial site. In *Avaiable online at http://www.ieeta.pt/robocup/*.