Design of an Expert System Coach for Complex Team Sports

Brice A. Colcombe, Lindsay M. Horton, Muhammad Ommer, and Julia Teng *Member, IEEE*

***Abstract*— The purpose of this paper is the creation of an expert system coach for complex team sports. As soccer analytics becomes a more advanced technological industry, a market has emerged to take soccer game data and create an automated system which is able to recognize complex patterns that coaches are unable to see and output tactical adjustments to increase the probability of winning. This system will use coaching input based on coaching strategies and formations. Full games can be simulated to test and validate that the system will increase the probability of winning more games in order to get to the NCAA tournament.**

CONTEXT

A. Soccer

The National Collegiate Athletic Association (NCAA) is the governing body that regulated specific set of rules in collegiate levels of sports [6]. Soccer is a game played with two opposing teams. Each team consists of ten field players and one goalie. The objective of the game is to score more goals than the opposing teams. A regular game is 90 minute long, the team with the most goals wins. If there is a tie then will be two 10 minute halves and the first team to score a goal will win the match. Traditionally, the way soccer has been played was heavily based on the coach's expertise and experience with the sport. They are responsible in making decisions during matches. However, soccer has slowly been making the transition in utilizing data analysis for making important decisions, which will be our focus in developing our Expert Coaching System.

B. *Netcentricity*

Soccer analytics only made significant progress in recent years, due to the complexity of the sport. An example showing soccer's complexity is the network diagrams in Figure 1. The diagram visualizes a European championship matches held in n 2008 between Spain and Italy. Each network illustrate passes made between team members in certain areas of the field. The numbered circular nodes indicate the player's jersey. The nodes are color-coded and sized to indicate the passing accuracy and player performance respectively. The arcs in between each node represent the passes made from one player to another. The width of the arc shows the successful passes made, the arc grows at an exponential rate based on the successful passes completed between 2 players. This visual network shows which player contributes the least or most, which can help coaches, make decisions for future matches [2].

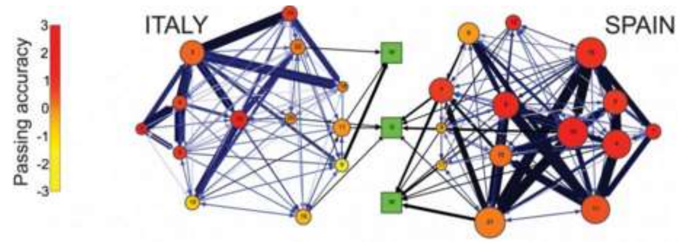

Figure 1: Network diagrams for Spain and Italy

II. ROLE OF A COACH

Coaches face many pressures and are constantly under scrutiny to perform and get results on the field. In order to be successful, a coach must be able to positively influence their players to commit to their coaching strategies and make any adjustments accordingly in order to follow the coaches' ideal tactics. Coaches often rely on their expertise to spot any cases where an adjustment must be made [3]. This is done in training and games. Coaches who only rely on this strategy of coaching are known as more traditional coaches and their current operations are modeled below in red. With many new technologies out to aid coaches, lots of coaches are turning to data driven coaching. This type of coaching uses their expertise as well as data collected during training and games to make any appropriate tactical adjustments. This is modeled below in blue.

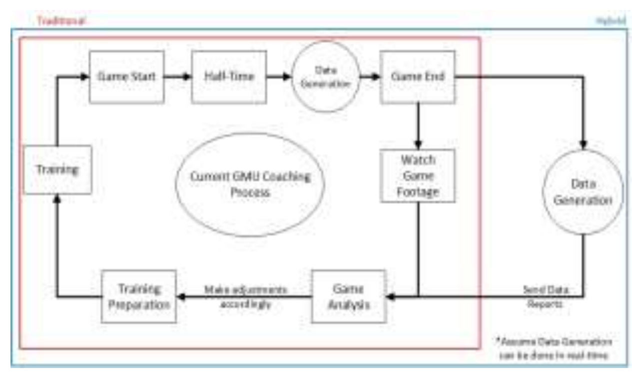

Figure 2: How coaching is done

III. STRATEGIES

In order to follow with their set game plan, coaches have a variety of different strategies which they hope to impose on their teams' in order to win games. These strategies often involve styles of play and formations that go along with them. Formations are how teams are set up on the field in order to have numerical advantages at certain places on the field. Some more popular formations include the 4-4-2 and the 4-3-3, both shown below in Figure 3 [5]. These formations both have different advantages and disadvantages at different areas of the field based on their numerical advantages. For example, a flat 4-4-2 has a lot of numbers at wide flank positions but may be outnumbered in the central midfield [5]. On the contrary, a 4-3-3 has a numerical advantage in central midfield but can be outnumbered at wide positions. With each formation, there are many variations that arise based on the style the coach implements. With a 4-4-2, coaches can play with a flat 4 midfield or a diamond midfield to have a numerical advantage in the central midfield. Small variations and changes are often made in game based on what a coach may see is a strength or weakness [5].

4-4-2 formation 4-3-3 formation

Figure 3: Two types of formation

STAKEHOLDER ANALYSIS

I. *Stakeholders*

A. GMU Men's Soccer Team Coaches (Head/Assistant)

The role the coach's is substantial as many of the decision-making criteria's fall upon the coaching staff. The coaches are funded by Georges Mason University (specifically the athletic department) and they train their soccer players in order to promote stronger performances in key matches. Ultimately, both head and assistant coaches want to develop the strongest possible soccer team formations for winning matches. Currently, GMU coaches rely on a hybrid strategy of traditional coaching formations as well as the use of data analytics reports provided by InStat. At collegiate levels, this is done in order to receive an NCAA bid at the end of the season to participate in the NCAA tournament [3].

B. GMU Men's Soccer Team Players

This includes all the current member of the George Mason University's Men's Soccer Team. The players are trained by the coaching staff during the regular playing season in order to participate in home and away matches. Their objective is to win matches consistently in order to increase their overall Rating Percentage Index (RPI) score for the respective season [4**].** The players traditionally need to have a strong connection and trusts with one another in order accomplish their objective of winning.

C. Trainers

The trainers work in conjunction with George Mason University's Athletic Department and their mission is to provide health rehabilitation services for players if they are fatigued or injured. This is determined by either the coaching staff or the players themselves who are sent to visit the training staff.

D. National Collegiate Athletic Association (NCAA)

The NCAA is a non-profit organization established in 1906 that promotes and regulates the athletic programs of colleges and universities [6]. This is done in order to maintain a set of standards that GMU Men's soccer team must abide by in order to promote fair gameplay. With these standard regulations come consequences derived from the NCAA. If college teams fail to meet standards, it can result in said consequences. As such, GMU is partnered with the NCAA committee to ensure that all teams have equal opportunities to succeed in their perspective sports.

E. Team Investors (George Mason University)

George Mason University works concurrently with the NCAA in order to establish universal rules and regulations that collegiate teams must abide by. These rules regulate what coaches can and cannot incorporate in their respective training programs. GMU also provides the athletic department with certain resources (i.e. scheduled training sessions, practice fields, equipment and salary) in the hopes that coaches will be able to develop a winning team, thus increasing the school's reputation in its respective athletic program.

F. Parents

The parents are the team players initial supporters as they have invested their own resources (money and time) in the players earlier years of their career. This instills the parents with a certain sense of pride as the players advance in soccer team rankings and hopefully become Division I athletes.

G. Academies

Academies are youth clubs that develop players before they become student-athletes in college sports. The players that wish to pursue college level sports can play for respective academies, thus improving their skills necessary for higher-level gameplay. Academies offer a variety of benefits for students such as scholarship opportunities.

H. Sports Analytics Company

InStat is a sport's data analysis provider that is officially partnered with GMU's athletic department. InStat provides statistical reports of previously played game data between GMU Men's soccer team and opponents. These reports are comprehensive (providing over 1400 statistics for one game), and the coaching staff utilizes in order to make strategic decisions in order to increase the probability of winning.

II. *Stakeholder Tensions*

Certain coaching disagreements arise between the head and assistant coaches who have their respective styles in how to structure a team to increase winning probability. With this, the assistant coach may not have much say in how to organize an effective team. The coaching staff is also pressured by the University to produce a winning team but is restricted to a budget if school revenue does not increase. Players are pressured by coaches and sometimes are not chosen to play in key matches, which inhibit team development. Trainers have little to no control on the level of intensity that is pushed onto players by the coaches which can result in future injuries.

The Expert Coaching system (ECS) is a strategy that will alleviate some if not all of the tensions present between our primary stakeholders. With this tool, the objectives of each stakeholder will merge in order to produce win-win solutions. Combining the ECS with GMU's current coaching strategy will enable stronger use of sports data analysis, which will enable in-depth understanding of soccer's complexity. Stronger understanding is key for coaches and players in understanding the team's strength and liability areas. This provides a way for the University to increase revenue and reputation as the team will be restructured in order to improve liability areas, thus increasing the probability of winning matches. Team restructuring will also prevent future injuries for trainers as well as increase player development as actual teammates.

GAP ANALYSIS

I. Atlantic 10 Conference History

George Mason is a member of the 13 school Atlantic 10 Conference. Each year, the Atlantic 10 holds a conference tournament [1]. The winner of this tournament gets an automatic bid into the NCAA tournament. Historically, Saint Louis is consistently the best soccer team in the conference. George Mason sits in the middle at 5th in terms of winning percentage over the past 10 years [8]7.

II. NCAA Tournament

The goal of every collegiate soccer team is to make the NCAA tournament. This tournament is a single elimination tournament featuring 48 teams [6]. 22 of the teams are the winners of their conference tournament while the remaining 26 teams are decided based on the teams' Rating Percentage Index or RPI. The RPI is for each team is calculated using a formula which takes into consideration 3 parts: a team's winning percentage, their opponent's winning percentage, and their opponent's winning percentage [4]. This shows the importance of a strong strength of schedule when deciding which teams to play every year.

III. Money

Coaches who consistently win their conference championship, make it to the NCAA tournament, and win the NCAA championships are substantially rewarded for their efforts through their salaries. Comparing the head coaching salary of George Mason to that of a consistently top team, there is a gap of about $67,457 per year [7, 9-15]. This number is based on both experience and performance at the respective schools.

IV. Gap

There is a gap between George Mason University Soccer and perennial Atlantic 10 powerhouse Saint Louis University's Soccer Team [11]. To replicate Saint Louis's success over the past ten years, George Mason will need to win two out of the every five years Atlantic 10 championship. They will also need to receive an NCAA bid six out of every ten years. To accomplish this, they will need to have an average RPI score of at least .56 [9-15]. By closing the gap between teams winning percentages, the head coach will be able to close the $67,000 salary gap.

PROBLEM AND NEED STATEMENTS

I. Problem Statement

George Mason University Men's Soccer Team over the past ten years has not been able to consistently receive an NCAA Tournament bid at a high rate; only two out of the past ten years [11].

II. Need Statement

There is a need for a tool that combines coaching expertise and soccer game data that will help coaches and players understand the complexity of the game. This understanding will lead to a competitive advantage over other teams which will increase the team's probability of receiving and NCAA Tournament bid three out of every five years (60% of the total bids).

CONCEPT OF OPERATIONS

Figure 4 shows the proposed concept of operations for the expert coaching system. This system will be implemented both at halftime and after the game once data is generated. The system will input the statistics and probability maps from the data generation and will be able to recognize patterns and process statistical data in order to output tactical adjustments that coaches are able to use in preparation for the second half as well as in preparation for any other upcoming games. This system will be able to 'see' the complexity that is often missed by the coaches, giving them a heavy advantage over their opposition.

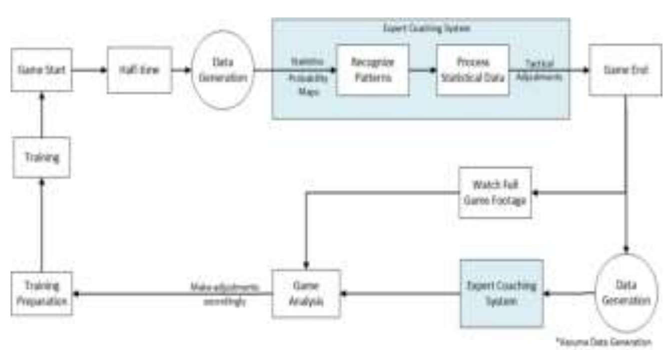

Figure 4: Proposed Concept of Operations

SIMULATION

The expert coach simulation will simulate a ninety minute soccer game based on previous game data. The code is based off of the previous senior design project, "Assessment of Soccer Referee Proficiency [16]. The code has been reverse engineered to its base simulation interface and then we have been able to adapt it to the expert coaching systems objectives. The simulation will run for the first half (forty five minutes) before running through the expert coaching rules to see if a tactical adjustment are necessary. The tool is designed to be used by the coaching staff in order to help them recognize difficult patterns that occur during the game.

II. Probability Maps

The input into the simulation is a set of fifteen probability maps that combine to make one strategy. The probabilities are formed based on previous soccer game data. There are fourteen zones per map. For each of the zones there is a separate probability map depending which zone the ball is in. The fifteenth probability map is the probability a shot will be successful based on what zone the shot was taken from. The set of maps will change completely from one strategy to another.

III. Ball Movement

The ball will always go through the same process in each zone. Once a ball has been successful received in a zone the player will have a choice to either pass the ball to a new zone or dribble the ball within the current zone. If they decide to dribble the ball then they must pass it after 5 seconds. Once the ball has been passed the ball will go through a transition. This transition period is when a ball can be turned over to another team or successfully passed to another zone. If a pass was a shot then the ball will automatically be turned over the the other team. This transition period is also when a new probability map will be called.

IV. Coaching Rules

The coaching rules were derived from interview with the George Mason University coaching staff. They follow the rules of each of the strategies. Since we are unable to model the players the field has been broken up into sections; left wing zone, right wing zone, middle zone, defensive zone, midfield zone, and forward zones. There is a goal total passing percentage of the total passes for each of the zone and all of the rules reflect that. If the rule is not being met then the coaching rule will tell what strategy to play to help give the team a competitive advantage.

V. Validation of Concept of Operations

Figure 5 shows how we will validate our simulation and coaching rules. The simulation will run starting with a set of probability maps (a strategy) until a half time report is produced. The data gathered during the first half of the game will then be run through the expert coaching rules and then run the second half of the simulation with a new set of probability maps. The same first half game data will also be run through the second half of the game without making any changes. Both outcomes of the games will produce a final report and then we will compare the results.

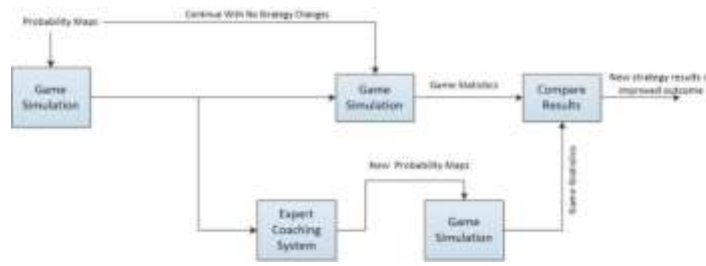

Figure 5: Process of validation

Once we run the simulation through a Monte Carlo simulation 1000 times we are expecting to see a shift from a very low probability of winning to a high probability of winning.

DESIGN OF EXPERIMENT

In order to simulate a variety of different strategies and formations, a design of experiment has been developed including many different formations. The following tables show the experiment which will be run to ensure the system can run all the scenarios. The tables break the field up into right, left, and central zones as well as the defensive, midfield, and forward zones.

Formations		Replications	Defense Zone	Midfield Zone	Forward Zone
4-4-2	Flat	250	≥ 20%	≤ 50%	≤ 30%
			< 20%	> 50%	> 30%
	Diamond	250	≥ 20%	≤ 55%	≤ 25%
			< 20%	> 55%	> 25%
4-3-3	Flat	250	≥ 20%	≤ 50%	≤ 30%
			< 20%	> 50%	> 30%
	Triangle	250	≥ 25%	≤ 55%	≤ 20%
			< 25%	> 55%	> 20%
Formations		Replications	Right Wing Zone	Left Wing Zone	Middle Zone
4-4-2	Flat	250	≤ 30%	≤ 30%	≤ 40%
			> 30%	> 30%	> 40%
	Diamond	250	≤ 25%	≤ 25%	≤ 50%
			> 25%	> 25%	> 50%
4-3-3	Flat	250	≤ 40%	≤ 40%	≤ 20%
			> 40%	> 40%	> 20%
	Triangle	250	≤ 30%	≤ 30%	≤ 40%
			> 30%	> 30%	> 40%

Table 1: Design of experiment table

Results

To come in final report

Recommendation

To come in final report

Acknowledgment

We would like to thank the Nathan Jones et al. who authored the Assessment of Soccer Referee Proficiency in Time-Sensitive Decision-Making project for their simulation to act as a base for ours. We would also like to thank Tom Morrell for the various simulation assistance.

References

[1] Atlantic10.com, Atlantic 10 Conference Official Athletic Site', 2015. [Online]. Available: http://www.atlantic10.com/SportSelect.dbml?DB_OEM_ID=3160&SPID=136213&SPSID=799925. [Accessed: 20- Oct- 2015]. [Accessed: 07- Dec- 2015]

We would like to thank the Nathan Jones et al. who authored the Assessment of Soccer Referee Proficiency in

[2] Axonpotential.com, 'the New Soccer Metric "Flow Centrality Axon Sports', 2015. [Online]. Available: http://www.axonpotential.com/the-new-soccer-metric-flow centrality/. [Accessed: 07- Dec- 2015]

[3]J. Perez, 'Concepts and Coaching Guidelines', *U.S Soccer Curriculum*, 2015. [Online]. Available: http://resources.ussoccer.com/n7v8b8j3/cds/downloads/Part%202 20%20Concepts%20and%20Coaching%20Guidelines%20U.S.%2 Soccer%20Coaching%20Curriculum.pdf. [Accessed: 07- Dec2015]

[4] Sites.google.com, 'RPI: Formula - RPI for Division I Women's Soccer', 2015. [Online]. Available: https://sites.google.com/site/rpifordivisioniwomenssoccer/rpi formula. [Accessed: 07- Dec- 2015]

[5] Soccer-training-guide.com, 'The 4-3-3 is a Highly Popular Formation', 2015. [Online]. Available: http://www.soccer-training guide.com/4-3-3.html#.Vi1ul36rTIU. [Accessed: 08- Dec- 2015]

[6] NCAA.com, 'NCAA.com "The Official Website of NCAA Championships', 2015. [Online]. Available: http://www.ncaa.com/. [Accessed: 07- Dec- 2015]

[7] Virginiasports.com, 'VirginiaSports.com - University of Virginia Official Athletics Website - UVA Cavaliers Men's Soccer', 2015. [Online]. Available: http://www.virginiasports.com/sports/m soccer/archive/va-m-soccer-archive.html. [Accessed: 07- Dec 2015]

[8]'2009 Atlantic Coast Conference', 2015. [Online]. Available: http://grfx.cstv.com/photos/schools/acc/sports/m soccer/auto_pdf/acc-records.pdf. [Accessed: 07- Dec- 2015]

[9] Uclabruins.com, 'Men's Soccer - Schedule - UCLA Bruins Official Athletic Site | UCLABruins.com', 2015. [Online]. Availablehttp://www.uclabruins.com/SportSelect.dbml?SPSID=79841&SPID=126914&DB_OEM_ID=30500&Q_SEASON=2013. [Accessed: 07- Dec- 2015]

[10] Gomason.com, 'Men's Soccer - News - GoMason.com – The Official Athletic Site Of George Mason University', 2015. [Online]. Available:http://www.gomason.com/SportSelect.dbml?DB_OEM D=25200&SPID=80375&SPSID=606456&DB_OEM_ID=25200. [Accessed: 07- Dec- 2015]

[11] Gomason.com, 'Men's Soccer - News - GoMason.com – The Official Athletic Site Of George Mason University', 2015. [Online]. Available:http://www.gomason.com/SportSelect.dbml?DB_OEM D=25200&SPID=80375&SPSID=606456&DB_OEM_ID=25200. [Accessed: 07- Dec- 2015]

[12] Umassathletics.com, 'University of Massachusetts Official Athletic Site - Men's Soccer', 2015. [Online]. Available: http://www.umassathletics.com/sports/m-soccer/umas-m-soccer body.html. [Accessed: 07- Dec- 2015]

[13] Vcuathletics.com, 'VCU', 2015. [Online]. Available: http://vcuathletics.com/sports/msoc/index. [Accessed: 07- Dec- 2015]

[14] Iuhoosiers.com, 'Official Athletics Site of the Indiana Hoosiers', 2015. [Online]. Available: http://www.iuhoosiers.com/index.aspx?path=msoc. [Accessed: 07 Dec- 2015]

[15] Masslive.com, 'Massachusetts state employee salary database 2015', 2015. [Online]. Available: http://www.masslive.com/politics/index.ssf/2015/03/massachusett_state_employee_salary_database_2015.html. [Accessed: 07- Dec 2015]

[16]N. Jones, A. Cann, S. Almashhadi and H. Popal, 'Assessment of Soccer Referee Proficiency in Time-Sensitive Decision-Making', *George Mason University*, 2012. [Online]. Available: http://seor.gmu.edu/projects/Capstone_2012/DMTSO_Final_Paper.pdf. [Accessed: 06- Dec- 2015]