Predictive Analysis from Advanced Sports Data For Effective Decisions

Sowmya N S
PG Student
Dept. of CSE,CIT Gubbi Tumkur

Mr. Justin A Gopinath
Associate Professor
Dept. of CSE, CIT Gubbi Tumkur

Mr. Harish T A
Assistant professor
Dept. of CSE,CIT Gubbi Tumkur

***Abstract*-In sports science, the data collected from field sensors, wearable inertial sensors and social media i.e., twitter (live tweets) are used to derive the performance of the games. Sensors and Twitter provides huge amount of data. We are doing analysis on three data set i.e., field sensor data, inertial sensor data and twitter data. The historical data of field sensor and inertial sensor are collected for the performance analysis. The analysis of twitter data that is live tweets from the twitter users during the game or after the game. The three different data set analysis gives the three different patterns. Finally, the three different patterns are combined using a contrast set mining algorithm based on players and game history of performance to give one predicted solution.**

This paper describes the performance analysis of individual player or team in the games using contrast set mining algorithm that establishes the relationship between field sensor data, inertial sensor data and twitter data. Finally the predictive results will be obtained in the visual form.

***Keywords*-Contrast-set mining, data mining, inertial sensor data, field sensor data, twitter data, predictive analysis.**

I. INTRODUCTION

The Performance analysis can be defined as the analysis of data or information to aid in the acceleration of player performance [1]. The credentials of performance indicators, repeatable techniques for assemblage of data and the way in which the information is analyzed, are all key aspects to a successful performance analysis.

Predictive analysis is the preparation of mining information from existing data sets in order to conclude patterns and predict future outcomes and trends. Predictive analysis does not tell you what will occur in the future [2]. In predictive modelling, data is collected a statistical model is formulated, predictions are ready and the model is corroborated as the further data becomes accessible. Predictive analysis is applied to several research areas, including meteorology, security, disaster avoidance, genetics, performance analysis, economics and marketing. Fig 1 shows how the predictive analysis used in the analysis.

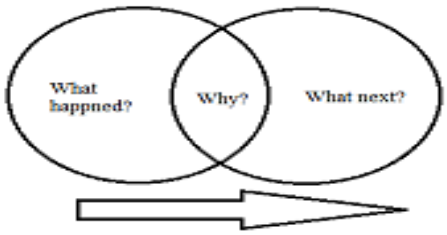

Fig1: Diagram of Predictive Analysis

Our system accumulates large volumes of data about a player as they record training, competition, tests, nutrition and health and wellness. It then uses classy quantitative methods to derive insights from this data, and assists coaches in drawing on those visions to shape decisions and, ultimately, improve performance. Instead of looking backward to analyze "What happened?" predictive analytics help coaches and players to answer the questions, "What's next?" and "What should we do about it?" [3].

Now a days, sports is one of the highly deliberated subject around the world. The media, fans, and the teams deliberate about the performance, mistakes and strategic decisions throughout the game or after the game. Some of the statistics about the sports are calculated manually by trainers and some further statistics are mined automatically by computer-based analysis systems. Such systems necessitate expensive computers to calculate the massive amount of data to process and some systems need too much time to calculate the statistics [4].

In any sporting circumstances, it is hard to notice and remember all the key events happening within a training period or match. Analysis based only on the precise observation and recall is a key

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

tool for taming future performance. Analysis is about making a valid and dependable record of performance by means of methodical observations. When we want to emphasis on player or game history performance, we should deliberate shooting the ball, getting the ball, scores of the each player etc., based on the shooting efficiency every miss costs a team about a point. A player's true shooting gives a pleasant measure of a player. Finally players nearer to the basket tend to get an edge, Position matters. If your player is a big man you will anticipate more efficient scoring and more plays that get them the ball. When you try and inspect if a player is good, it's not a bad idea to just add up how many points they get how many points they get, how many times they get the ball and subtract how many times they give up the ball, higher is obviously better [5].Many existing sports analysis systems are using advanced techniques such as field sensors, inertial sensors and social media (twitter).

Fig2: Team Training

Team Training curriculums are deliberate to deliver a challenging workout that is systematized, instructive, advanced, and modified for entire sport teams [6]. Fig 2 shows the team training in sports. Training sessions are classically accompanied at the team's training facility and emphasis primarily on linear speed and multi-directional skill development. Team Training curriculums are constructed to be sport pertinent with medicine ball strengthening, bodyweight repelled power improvement, energy system improvement, and speed ratios that smear directly to specific sports demands.

Sports science consuming advanced technologies such as inertial sensors, high speed cameras, eye tracking, social media, etc., [7] for the performance analysis. In this paper, we are allowing for the data sets of three technologies i.e., inertial sensor data, field sensor data and twitter data. Wearable inertial sensors are small in size used throughout the training period. Sensors are positionedon player's different parts of the body while coaching that affords enormous amount of data for each motion of the player in the game. The field sensors i.e., several cameras located on the field, that captures the sports events take place on the playing field. It affords enormous amount of data for every activities on the field. Twitter is the major issue in sports data analysis for the publics, professionals, team, and player's to do conversation about the performance of the games by means of tweets. Gigantic number of tweets may come throughout the game or after the game and it will be amplified. The tweets may be positive, negative or neutral established on the game performance. The tweets from the twitter users are beneficial for the analysis of each individual player's or the team and for choosing the players in the further games. Consequently all these three technologies produces enormous amount of data.

As said above, the technologies produces enormous amount of data, subsequently the analysis of that data based on performance is less and also the analysis of collective features of dissimilar data sets is less. We are using contrast set mining algorithm for the collective feature analysis of three data sets (inertial sensor data, field sensor data and twitter data). Contrast set mining algorithm establishes the relationship between those three data sets.

In this paper, the use of contrast set mining algorithm to afford the collective predictive analysis of three data sets [8]. Thus one predicted solution for the three data sets has been provided in the visual form.

II. RELATED WORK

In the existing system of using the sports advanced technologies produces enormous amount of data for the analysis of game performance. The technologies produces enormous data and instantaneously monitors the team players throughout the game or monitors after the data analysis. But the possibility of analyzing the enormous amount of advanced sports data is less and the analysis of collective data set is correspondingly less in scope. Some of the complications in the analysis of game performances are below:

A. Less Scope on Opinion Mining:

Sentiment analysis or opinion mining is the field of study that analyses people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as player performance, team performance, and tactical decisions throughout game in sports. People thoughts and opinions are shared prolifically through online

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

social networks such as twitter. To make use of this information, we need to be able to distinguish between what important and interesting is related to game [9]. The people's opinions about sports are used to recoverperformance of player's or team or for selecting a player for further games are less used.

B. Poor utilization of sensor provided data:

Wearable sensors are small in size, placed on the players body throughout training period to get the data when the players are batting, jumping, throwing, etc.,. Sensors measures characteristics of player's activities and convert them into information that may use or kept digitally for later access and analysis. Sensors collects enormous amount of data from the players activities in the game and have a primary role in subsequent processing and transport of data [10]. The analysis on sensors provided data in sports are less.

Sensors produces enormous amount of data at a time for each activity of the players. Monitoring is a process that observes a state in time or tracks changes in data sets to derive information. We don't know the collected data from sensors are correct or not. Without knowing about the data collected from sensors in the training period of players, we can't able to monitor players at a time when playing the game.

C. Different data mining algorithms concentrating only on one area of data set:

Currently the analysis of the game performance is centered only on the one data set. The possibilities of collective features analysis of different data sets are less in scope. For example the analysis of game performance by means of social media such as Twitter, Facebook, etc., separate analysis for each individual data set will be time consuming.

III. PROBLEM STATEMENT

Through the advanced technologies like wearable sensors, field sensors, social media, etc., in the sports science aimed at the analysis of performance of the certain games. Those advanced technologies generates enormous amount of data. Most of the study is based on the player's or game history of performance. Wearable sensors used merely throughout the training period and it delivers enormous amount of data for each player activities. A Field sensor like cameras positioned around the playing field and captures the every movement on the ground and delivers huge amount of data. Social media such as twitter acquires enormous amount of tweets may be positive, negative and neutral throughout the game or after the game. The study of consuming these enormous amounts of data is less. The study remains only for one dataset.

Hence to ensure the predictive analysis via merging the features of three different datasets i.e., inertial sensor data, field sensor data and twitter data. We get the predictive results.

IV. IMPLEMENTATION

A. Proposed System

In the proposed system we are carrying out predictive analysis of different or several factors related to game performance in sports on the source of both historical and live data. We also shown the establishment of relationships among different data's to perform a predictive analysis.

Advantages:

- Wearable sensors do not require a dedicated space for experiments.
- Potentially capable of helping us to evaluate the performance of players.
- Player/team can use social media (twitter) to directly connect with their fans to get feedback.
- Logical analysis of the data improves performance.
- Collective predictive analysis will be helpful for players selection in future games.

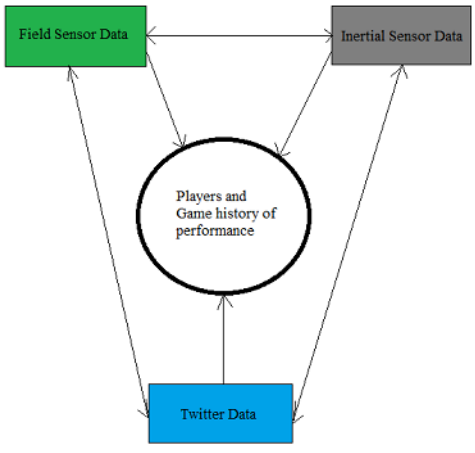

Fig3: Architecture diagram of proposed system

Fig 3 shows the system architecture that has three modules field sensor data, inertial sensor data, and twitter data.The analysis of sensor provided data are enormous amount of data that is quite challenging

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

task. A Field sensor captures the sports activities that occur on the playing field throughout the game. It delivers the data while game is on. Wearable inertial sensors are used for sensing and monitoring the individual players throughout training period. Sensors are placed on players body for the automatic recognition of player's each activity. It provides data for the players activities like running on the treadmill, exercising on stepper and playing the basketball.Now a day, Twitter is very popular and it is beneficial for prediction based on opinion in their tweets.In twitter millions of users share their opinions in short messages called tweets on sports.

Since all modules are collectively combined by establishing mutual relationship based on the players and game history of performance, by this combined analysis we can easily predict the future game performance and players selection.

B. *Methodology*

Detecting patterns of factors associated with game performance is of high interest to the sports community. We applied the Contrast Set Mining algorithm to analyze sensor data and twitter data in contrast to the players and game history of performance. The performance database results were compared.

Phase 1: Field Sensor Data

Field sensors i.e., cameras, captures the sports accomplishments as they happen on the playing field or in a laboratory setting. In the field, whether it is used as a coaching tool or for study purpose and camera system to quickly scrutinize motion. Many accomplishments in sports require the motion to be captured with high speed cameras so as to capture the most precise aspect of every movement. Camera movements can also be motion bagged so that a virtual camera in the field will pan, tilt or dolly around the field determined by a camera operator while playing the game, and the motion capture system can capture the camera and objects (ball, mesh, bat, etc.,) as well as the players performance. This permits the computer-generated characters, images and sets to have the same perspective as the video images from the camera. A computer processes the data and shows the movements of the players, providing the preferred camera positions in terms of objects in the set[7] [11]. The analysis of this data will provide each team performance or player performance and each object movements in the field.

Phase 2: Inertial Sensor Data

Miniscule inertial sensor units that encompass accelerometers and gyroscopes are occasionally accompanied by magnetometers. The size, weight and cost of wearable sensors are less. The main advantages of inertial sensors are that they are self-contained, non-radiating, non-jammable devices that provide dynamic motion information through direct measurements in 3D. A recent application domain of inertial sensing is automatic recognition and monitoring of human activities [12].

Fig4: Miniature Wearable Inertial Sensor

The fig 4 shows the wearable inertial sensor which is used to put on the player's body. They are used in sports during training period mainly for improving the performance of players in the game. It senses and collects the huge amount of data for each activity of the player and monitors at a time [13]. The logical analysis is based on three activities i.e. running on treadmill, exercising on stepper and playing basketball.

Phase 3: Twitter Data

Microblogging today has become a very popular communication tool among internet users. Millions of users share opinions on sports during game or after the game every day [14]. In this paper we focus on using twitter, the most popular microblogging platform for the task of sentiment analysis. We collect live tweets from the twitter on the game performance or about the tactical decisions taken during the game. In twitter, the tweets on any particular game in sports may come from common people and from the experts in that game. Users tweet about the team performance, individual player performance, scores in the game, and tactical decisions from the umpire during the game in very short messages. The figure shows example of typical posts from twitter about sports.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

Fig5: Example tweets on IPL match

Fig 5 shows the example of tweets from twitter on the IPL match between Delhi Daredevills and Rajasthan royals. The tweets from the common people and experts in the particular game upsurges each time. As the spectators of social media services grows every day. Data from these sources can be used in opinion mining and sentiment analysis tasks. For example, the sports team or player can directly connect with their fans, may be interested in following questions:

- What do people think about the game performance (player, team)?
- How positive or negative are people or experts about the performance?
- What would people or expert prefer about player or team performance to be like?

We do three activities before classifying the tweets into positive, negative and neutral.

- **Display the tweets:** the live tweets such as previously stored tweets and upcoming tweets are displayed. The tweets from people increases until we stop it.
- **Save the tweets:** In the first activity, it displays only the tweets but it is not saved. Here it saves the past and present tweets and it increases until we stop it.
- **Clean the tweets:** the tweets saved in the second activity are used for cleaning the tweets. It means clean the tweets in order.

We collected huge amount of text posts from twitter. We are taking only expert tweets for the analysis [15]. We classify the tweets into three sets of texts:

- Texts containing positive opinion on performance, such as good, well, great, etc.,
- Texts containing negative opinion on performance, such as bad, poor, very bad, etc.,
- Objective texts that only state a fact or do not express any emotions.

Data: The data used in the study consists of inertial sensor data, field sensor data and Twitter data combined to game performance.

- Inertial sensor data collected from UCI Machine Learning Repository: Daily and Sports Activities Data Set [16].
- Field sensor data collected from advanced technologies in sports (high speed cameras).
- Real time twitter contains data of social media

Each report in these databases consists of structured fields plus an unstructured Narrative explaining the event.

Problem Definition: In association rules, we classically deal with sports data where the database D is a set of transactions with each transaction T, I = {I1,I2,I3,I4,,,,,,,,,,,,,Im}. Each affiliate of I is a literal called an item, and any set of these literals is called an item set. We simplify the data model to gathered categorical data. The data is a set of k-dimensional vectors where each element can take on a predictable number of discrete values. The vectors are systematized into n mutually elite groups [17].

The concept of an item set can be extended to a contrast-set as follows:-

Let A1,, A2,,,,,,,,,Ak be a set of k variables called attributes. Each Ai can take on values from the set {Vi1, Vi2, and Vim}. Then a contrast-set is a conjunction of attribute-value pairs amorphous on groups G1, G2,,,,,,,,,,,,,Gn.

Example: (game = basketball) ^ (result= win).

We define the support of a contrast-set with respect to a group ‘G ‘as follows

Our goal is to find all contrast-sets whose support differs meaningfully across groups. Formally, we want to find those contrast-sets

For all ijP(cset = True | Gi) != P(set = True | Gj)…… (1)

We call contrast-sets where Equation 1 is statistically valid significant.

A mining algorithm: We treat the problem of mining contrast-sets as a tree search problem. The root node is an empty contrast-set, and we make children of a node by concentrating set by adding one more term.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

We use a canonical ordering of attributes to evade visiting the same node twice. Children are shaped by appending terms that follow all existing terms in a given ordering

We search this tree in a breadth-first, level wise manner. Given all nodes at a level, we scan the database and count their support for each group and then examine each node to determine if it is significant.

Steps:

- Plot a diverse graph by considering different attributes for each module
- Extract the highly influential factor individually from each module
- Institute the connection between any two item sets
- Merging all item sets collectively and get mandatory pattern
- Prune out residual unmatched patterns and save it for future use
- Continue these steps until pattern represents in visualization.

Given all nodes at a level, we scan the database and count their support for each group and then examine each node to determine if it is significant and large, if it should be pruned, and if children should be generated.

After finding all substantial contrast-sets in the data, we then process the results and select a subset to show to the user. We show the low order results first, which are simpler, and then show only the higher order results that are surprising and considerably different.

Finding Significant Contrast Sets: We can plaid if a contrast-set is substantial by testing the null hypothesis that contrast-set support is equal across all groups or, alternatively, contrast-set support is independent of group membership.

The support counts from each group are a method of frequency data which can be analysed in contingency tables and pictorially characterize by plotting the graph. We form a 2*c contingency table where the row variable represents the truth of the contrast-set and the column variable indicates the group membership.

Algorithm:

Search and Testing for Understandable Consistent Contrasts.

Input: data D
Output: Dsurprising
Begin
Set of Candidates C {}
Set of Deviations D {}
Set of Pruned Candidates P {}
Let prune(c) return true if c should be pruned

1. while C is not empty
2. scan data and count support ¥ c € C
3. for each c € C
4. if significant(c) ^ large(c) then
5. D ← D U c
6. if prune(c) is true then
7. P ←P U c
8. else Cnew ←Cnew U GenChildren(c,P)
9. C ← Cnew
10. Dsurprising ← FindSurprising(D)

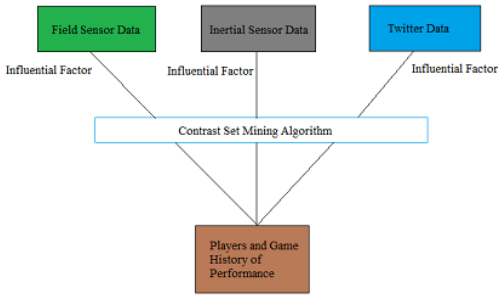

Fig6: Diagram shows how algorithm applied on modules

In the fig 6, contrast set mining algorithm applied on the three modules based on the players and game history of performance in the analysis.

V. RESULT ANALYSIS

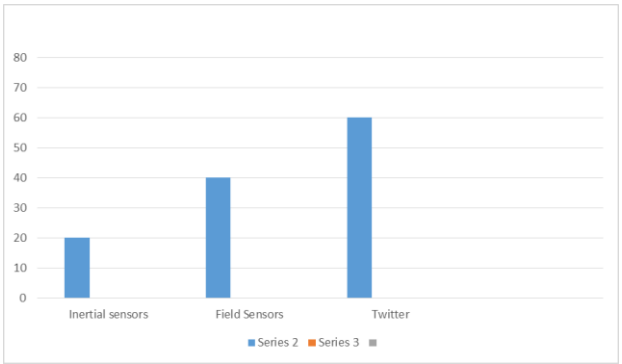

Fig7: Graph of highly influenced factor in performance analysis

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET

The fig 7 shows the highly influenced factor that is obtained from the overall combined predictive analysis using contrast set mining algorithm and as per result analysis the highly influenced factor is on Twitter data.

VI. CONCLUSION

This paper summarized that performance analysis of the games and players will be easy by combining the features of inertial sensor data, field sensor data and twitter data based on players and game history of performance. By establishing the relation between the datasets will shows the highly influential factor among all the dataset in the analysis of performance. This will be helpful for selecting the players or team in further games. The collective analysis also be useful for improving the performance in the games.

ACKNOWLEDGEMENTS

We are grateful to express sincere thanks to our faculties who gave support and special thanks to our department for providing facilities that were offered to us for carrying out this project.

REFERENCES

[1]Pro Stergiou (M.Sc.) & Larry Katz (Ph.D.).http://www.ucalgary.ca/strl/research/pas, PERFORMANCE ANALYSIS IN SPORT.

[2]www.webopedia.com/TERM/P/predictive_analytics.html, Predictive Analytics.

[3]http://www.metrifit.com/health-andsporttechnologies/predictive-analysis/#.VWCT5kZjKe0, Predictive Analysis.

[4] Michael Brand, Tobias Brandt, CarstenCordes, Marc Wilken, TimoMichelsen. Herakles: A System for Sensor-Based Live Sport Analysis using Private Peer-to-Peer Networks, Escherweg 2, 26129 Oldenburg, Germany.

[5]http://www.pponline.co.uk/encyc/how-performance-analysis-can-improve-your-coaching-methods-39, How performance analysis can improve your coaching methods.

[6]http://sportsnscience.utah.edu/team-psychology-technical/The Psychology of Individual and Team Sports (Technical), May 11, 2014November 21, 2014 by SportsnScience Writer.

[7]http://www.mi-as.com/applications/sport-analysis-and-performance/ Motion analysis software for sport performance.

[8]ZohrehNazeri, Daniel Barbara, Kenneth De Jong,George Donohue, and Lance Sherry. Contrast-Set Mining of Aircraft Accidents and Incidents, P. Perner (Ed.): ICDM 2008, LNAI 5077, pp. 313–322, 2008. © Springer-Verlag Berlin Heidelberg 2008.

[9] Bing Liu, OPINION MINING, Chicago, IL 60607-0753 liub@cs.uic.edu.

[10]Wixted, Andrew James, D. C. Billing, and Daniel Arthur James. "Validationof trunk mounted inertial sensors foranalysing running biomechanics under field conditions, using synchronously collected foot contact data." *Sports Engineering* 12.4 (2010): 207-212.

[11]http://www.xcitex.com/sports-performance-motion-capture-analysis-software-data-video-systems-athletic-applications-xcitex.php, Motion Capture.

[12] BillurBarshan∗and Murat CihanYüksek, Recognizing Daily and SportsActivities in Two Open Source Machine Learning Environments Using Body-Worn Sensor Units, The Computer Journal, Vol. 57 No. 11, 2014.

[13]http://www.acma.gov.au/webwr/_assets/main/lib310665/sensing-and-monitoring.pdf, Sensing and monitoring Recent developments SEPTEMBER 2011.

[14] Alexander Pak, Patrick Paroubek, Twitter as a Corpus for Sentiment Analysis and Opinion Mining, Universit´e de Paris-Sud, Laboratoire LIMSI-CNRS, Bˆatiment 508,F-91405 OrsayCedex, France.

[15] Smitashree Choudhury1, John G. Breslin2, Extracting Semantic Entities and Events from SportsTweets, #MSM2011 _ 1st Workshop on Making Sense of Microposts.

[16]ShrutiJauhari, AniketMorankar, Ernest Fokoue,http://archive.ics.uci.edu/ml/datasets, UCI Machine Learning Repository- center for machine learning and intelligent systems, 2013.

[17] Stephen D. Bay and Michael J. Pazzani, Detecting Change in Categorical Data: Mining Contrast Sets,We used C. Borgelt's implementation of Apriori version 2.1.

First Author: Sowmya N S was born in Karnataka, India. She received the B.E Degree in Information Science and Engineering from Visvesvaraya Technological University Belgaum Dist., India in 2013 and Mtech Degree in Computer Science and Engineering from the same University. Her research interests are in the area of Data Analysis and Predictive Analytics.

Second Author: Justin A Gopinath M-tech, was born in Karnataka, India. He is working as Associate Prof. inComputer Science Engineering Dept., CIT, Gubbi, Tumkur (district), and Karnataka, INDIA.

Third Author: Harish T A M-tech, was born in Karnataka, India. He is working as Asst.Prof. In Computer Science Engineering Dept., CIT, Gubbi, Tumkur (district), Karnataka, INDIA.

ISSN: 2278 – 1323 All Rights Reserved © 2015 IJARCET