Skillful Storyteller for Color Commentary in Sports

Ms. Amruta Waghulkar
Department of Computer Science and Engineering
Priyadarshani Institute of Engineering and Technology
Nagpur, *India*
e-mail: amrutawaghulkar91@gmail.com

Dr.A.R.Mahajan
Department of Computer Science and Engineering
Priyadarshani Institute of Engineering and Technology
Nagpur, *India*
e-mail: armahajan@rediffmail.com

***Abstract*—** The method of automated narrative is computerized sports interpretation. To uphold the member of the audience well-informed and amuse sports commentary is needed. By conveying long stories proper in the direction of the match in progress is one of the way through which one can amuse the onlookers. A system is presented that can indorse stories mechanically for interpreters to be told for the period of games. The stories would contain statistics from the preceding related games. This statistics would consist of data that match the present-day game statistics with the earlier game statistics, some connected facts about the players and teams involved. Commentating in sports generally involves two people – a play-by-play commentator and a color commentator. Play-by-play commentating comprises of transmitting to the onlookers what is actually happening in the pitch of play. Beyond reporting the activities of the players as they occur, the play-by-play commentator typically comments such facts as the score of the game, upcoming batters and statistics for the teams and players involved in the game. Color commentary, on the other hand, is much more subjective and wide, with the purpose being to increase entertainment (i.e, "color") to the broadcast. It will be a footstep in the direction of automate sports commentary and, thus, automate sequence of events.

Keywords: Automated Narrative, Information Retrival, Color Commentary.

*****

I. INTRODUCTION

Broadcasting of outdoor sports is an industry of billion dollars [1]. Transmission of most proficient sports to the public is thru TV set that ranges lots of home. The most significant part of the live watching experience is its commentary and it differs from the television experience. There has been much exploration done in the significance of commentary in the course of sports broadcasting. Maximum attention is given towards the words of commentator while seeing a game on television. The commentary ensures the influence of catching the attention of the onlooker to the parts of the picture that worth nearer attention [2], an effect called *italicizing* [3]. A mood can be set through commentary during broadcast. Generating a friendly atmosphere by commentator thru a broadcast makes the seeing experience more enjoyable. So as to listen to the analyses of the commentators fans often carry radios to live games as the explanations given in a broadcast by the commentators are suitable. Also, a handheld video device is supported by some sporting venues that provide the spectator with in-game commentary.

Serving the onlooker to monitor the game is the purpose of commentators and adding to its amusement value. To convey exciting, related stories from the sport's past is one way to add entertainment to a broadcast. A rich record is provided from the prolonged existence from which attractive stories could be strained. "Storytellers" are those who are popular commentators [8], as they enrich the games they call by adding stories.

A type of narrative discourse [7] is sports storytelling. Narrative discourse comprises of conveying to onlookers a series of actions in an exciting and engaging approach which is an inspired activity. Describing dependent events from the past with one or more main characters is the indication of narrative discourse. A challenging problem for Artificial Intelligence (AI) is automating narrative discourse and a subject of much recent research. The emphasis of this work is on one side of computational narrative that of selecting the narrative components based on the recent situation.

Sports story selection can be computerized with AI [4]. An AI system is developed to attain this objective that conveys stories in the framework of cricket. To place forward suitable stories to a (human) broadcast team this can be used all through cricket games that output a relevant story to the audience.

First, story-based sports commentary was formalized as a mathematical problem. Second, machine- learning methods and information retrieval techniques were used to solve the problem [5]. The solution could be for the specific domain of sports story selection, but also could be used in other domains involving story selection as the solution is enough general. Third, the approach was implemented in the domain of baseball; the resulting AI system was evaluated by observing feedback from human participants, and showed that it was effective in performing two separate tasks: 1) automating sports commentary, and thus automating narrative in a special case; and 2) assisting human commentators. The combination of information retrieval techniques was able to map previously unseen baseball game states to stories in a sufficiently effective manner improved the enjoyability of baseball broadcasts, increased the interest in watching baseball, and suggested stories to professional commentators that they would tell during a live game were shown.

II. RELATED WORK

StatSheet and Narrative Science [4] mechanically write previews for sports games that have not happened so far, and summaries about sports games that have previously occurred. For summaries, they are given with statistics from a accomplished game and create a narrative about the game, with the aim being to give an exciting summary of game events. For previews, they are given with statistics from previous games, and create a preview for the game that should attract the reader to lookout said game. Neither Statsheet nor Narrative Science function with live game information, both

are incapable to resolve the problem of delivering live stories through the game. They may, nevertheless, deliver other potential record of stories about previous games to expand our current system.

RobotWorld-Cup Soccer (RoboCup) is a test-bed for researchers, involving robots playing soccer [Kitano et al., 1997]. There is also a RoboCup simulation league, where the games, are not actually played, but are simulated on a computer. Earlier theoretical work in automated commentary has concentrated primarily on automated play-by-play commentary. Byrne, Rocco and MIKE [6] are three systems that create automated play-by-play commentary, for RoboCup simulator league games. The three systems acquire their information from the Soccer Server which recapitulates the gameplay's key features – the player positions and orientations, the score of the game. Each system produces natural language templates, putting in player and team names where suitable, and then uses text-to-speech software to express the resulting commentary. Dynamic Engaging Intelligent Reporter Agent (DEIRA) is alike to Byrne, Rocco, and MIKE, as it accomplishes the same chore, but in the game of horse racing. Rocco, MIKE, Byrne and DEIRA are intended to offer precise play-by-play commentary for RoboCup soccer and horse racing, and they achieve that closely. There are some efforts contained by these systems to deliver color commentary, but no one go to the extent that to try to include storytelling. That is, these systems hold a problem diverse from what this work tackles. The system could be used in combination with each of these automated play-by-play systems to generate fully automated commentary, containing both play-by-play and color.

Rocco (RoboCupCommentator) is template-based, so then when an incident is to be delivered to the onlookers, the suitable template is filled with the appropriate names of teams and players and then output. An instance of this would be a template " BLANK shoots and scores! BLANK now has the lead!" being filled with the player name "Smith" and the team name "New York" to develop" Smith shoots and scores! New York now has the lead!" Recurrence is avoided by keeping an account of previously spoken sentences. This also evades resaying a player's name too often, where a pronoun could be used in its place (i.e., "he" instead of "Smith"). Rocco retains a queue of likely outputs. If the system deems, it has nothing of importance in its queue, background information (such as the names of the teams involved in the game) is transported to the user in its place.

Byrne is alike to Rocco, but also contains a human face that alters its expression based on in-built biases and what is occurring on the computer-generated field. Where Byrne varies from Rocco is in its addition of emotion to the comments. The human face has static features for instance a nationality and approving or disapproving particular teams. These features are portion of the emotion-generation module used in Byrne to determine what emotions to assign to a given output. Each emotion holds a type (i.e., "sad"), a score (an intensity value for the emotion), a decay function (the emotion diminishes with time), a cause (what happened to cause the emotion) and a target (which is optional – where the emotion is directed).

MIKE (Multiagent Interactions Knowledgeably Explained) uses six Soccer Analyzer Modules that perform different actions on the available information, communicate with each other, and make proposals to the proposition pool, which results into natural language output. There are three lower-level modules – basic, shoot and technique and three higher level modules – bigram, Voronoi, and statistic.

III. PROPOSED METHODOLOY

In this section, we put forward an AI methodology to resolve the problem of conveying story-based color commentary to a live cricket game. We make use of three algorithms, first is the hierarchical clustering algorithm, second is the association rule algorithm and third is the pattern matching algorithm.

A. Hierarchical clusturing algorithm

Hierarchical clustering is an agglomerative (top down) clustering method. As its name proposes, the notion of this method is to construct a hierarchy of clusters, presentation relationships between the distinct members and merging clusters of data based on resemblance. In the first step of clustering is that the algorithm will search for the two most similar data points and unite them to produce a new "pseudo-data-point", which signifies the average of the two merged data-points. Each iterative step takes the following two closest data-points (or pseduo-data-points) and merges them. This procedure is usually repeated until there is one large cluster comprising all the original data-points. Hierarchical clustering results in a tree, presenting the relationship of all the original points. The tree is not a particular set of clusters, but rather a multilevel hierarchy, where clusters at one level are combined as clusters at the next level There are lots of times when clusters have sub-classes within them, and which in turn has sub-classes of their own. For example, in our system a cluster i.e. "cricket match" has sub-classes like the categories of different matches and in turn the matche has their own different sub-classes.

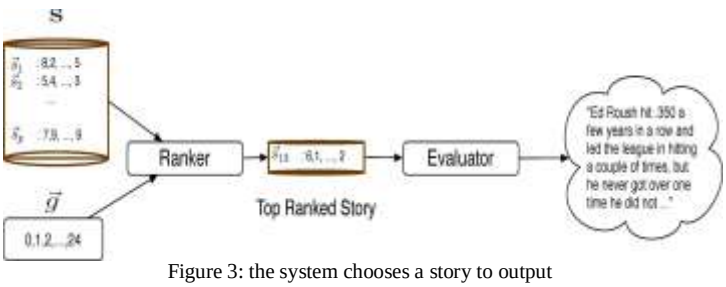

Figure 3: the system chooses a story to output

B. Association rule algorithm

There are numerous mining algorithms of association rules. One of the most popular algorithms is Apriori, which is used to mine frequent itemsets from large database and attaining the association rule for determining the knowledge.

1) Apriori algorithm :
Apriori algorithm [9] is simple to execute and very easy, it is used to mine all frequent itemsets in database. The algorithm sorts many searches in database to discover frequent itemsets where k-itemsets are used to create k+1-itemsets. Each k-itemset essentially be larger than or equal to

minimum support threshold to be frequency. Else, it is called candidate-itemsets. At first, the algorithm scan database to search frequency of 1-itemsets that holds only one item by counting each item in database. The frequency of 1-itemsets is used to search the itemsets in 2-itemsets which in turn is used to search 3-itemsets and so on till there are no more k-itemsets. If an itemset is not common, any large subset from it is also non-frequent. In the process of Apriori, the following definitions are needed:

Explanation 1: Suppose I={I1, I2, ⋯ , Im},(m⩾1) is a set of transactions, Ii= {T1, T2, ⋯ , Tn},(n⩾1) is the set of items, and k-itemset = {i1, i2, … , ik},(k ⩾1) is also the set of k items, and k-itemset ⊆ T.

Explanation 2: Suppose σ (item-set), is the support count of item-set or the frequency of existence of an item-set in transactions.

Explanation 3: Suppose Mk is the candidate item-set o size k, and Dk is the frequent item-set of size k.

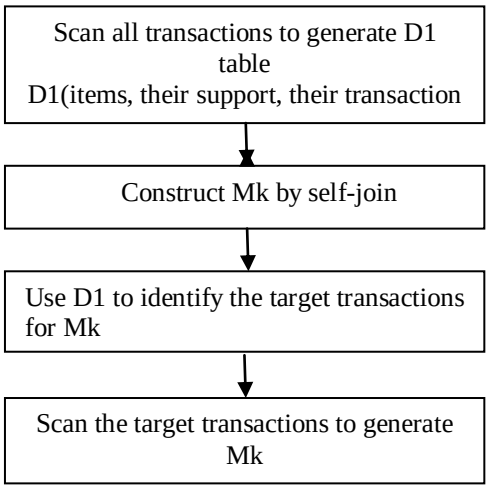

Figure 1: Steps for Ck generation

We utilize the Apriori algorithm to reduce the time consuming for candidates itemset generation. We initially scan all transactions to create D1 which comprises the items, their support count and Transaction ID where the items are found. And then we use D1 well ahead as a helper to generate D2, D3 ... Dk. At what time we want to create M2, we make a self-join L1 * L1 to construct 2-itemset M (x, y), where x and y are the items of M2. Formerly scanning all transaction records to count the support count of each candidate, use D1 to get the transaction IDs of the minimum support count between x and y, and thus scan for M2 only in these specific transactions. The same thing for M3, construct 3-itemset M (x, y, z), where x, y and z are the items of M3 and use D1 to get the transaction IDs of the minimum support count between x, y and z, then scan for M3 only in these specific transactions and repeat these steps until no new frequent item-sets are recognized. The entire method is shown in the Figure 1.

C. Pattern matching

Pattern matching [10] is to search a pattern, which is comparatively small, in a text, which is supposed to be very large. Patterns and texts can be 1-D, or 2-D. The existing pattern-matching algorithms match the pattern precisely or nearly within the text. An precise pattern-matching is to search all the occurrences of a particular pattern (x = x1 x2... xp) of p-characters in a text (y = y1 y2 ... yq) of q-characters which are assembled over a finite set of characters of an alphabet set denoted by Σ and the size of this set is equal to σ. The straight approach to this problem is to compare the first p-characters of the text and the pattern in some predefined order and, after a match or a mismatch, slide the entire pattern by one character in the forward direction of the text. This procedure is repetitive until the pattern is positioned at the (qp+1) position of the text. Pattern-matching algorithms test the text with the use of a window, whose size is equivalent to the length of the pattern. The initial step is to align the left ends of the window and the text and then relate the corresponding characters of the window and the pattern; this procedure is known as attempt. Subsequently a match or a mismatch of the pattern, the text window is moved to the right. The shift value differs based on the procedure used by different algorithms. This process is frequent until the right end of the window is within the right end of the text.

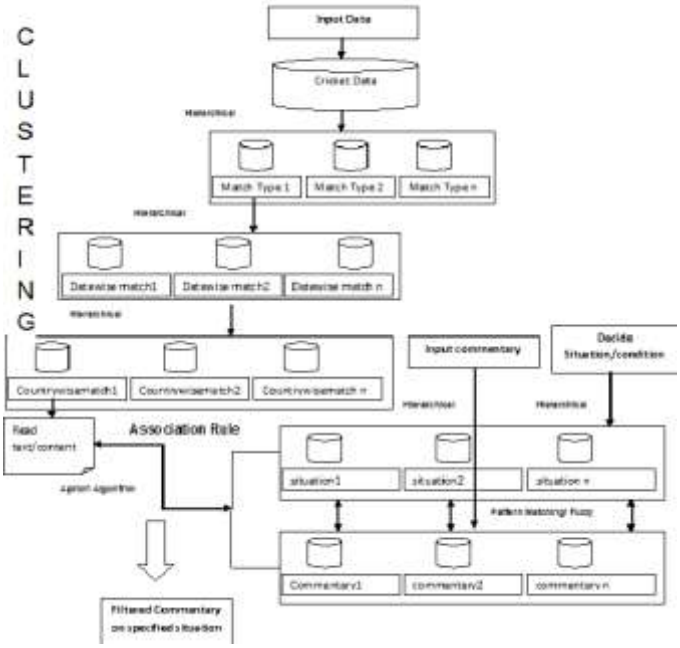

Figure 2: Proposed Architecture

IV. CONCLUSION

Storytelling is believed to be a cognitively rich and creative task. In order to excel in storytelling, an innate aptitude and training are required. In this work, we took a step towards automating this task by building the AI story selector for color commentary in any sport. This system can create a statistically significant positive influence on the sports viewing experience across several metrics.

V. ACKNOWLEDGMENT

I would like to thank my lecturers of the Department of Computer Science who helped me in my academic works and guided me throughout the session.

REFERENCES

[1] Greg Lee, Vadim Bulitko, and Elliot A. Ludvig" Automated Story Selection for Color Commentary in Sports" IEEE TRANSACTIONS ON COMPUTATIONAL INTELLIGENCE AND AI IN GAMES, VOL. 6, NO. 2, JUNE 2014.

[2] IGN, "MLB12: The Show Review," 2012 [Online]. Available: http://ca.ign.com/games/mlb-12-the-show/

[3] M. Kory, "Out of left field: Scouting the scout," *Baseball Prospectus*, May 7, 2012 [Online]. Available: http://www.baseballprospectus.com/article.php?articleid=16835

[4] G. Lee, "Automated story-based commentary for sports," Ph.D. dissertation,Dept. Comput. Sci., Univ. Alberta, Edmonton, AB, Canada,2012.

[5] G. Lee, V. Bulitko, and E. Ludvig, "Sports commentary recommendation system (SCoReS): Machine learning for automated narrative," in *Proc. Artif. Intell. Interactive Digit. Entertain. Conf.*, 2012, pp. 32–37.

[6] E. Andre, K. Binsted, K. Tanaka-Ishii, S. Luke, G. Herzog, and T. Rist, "Three RoboCup simulation league commentator systems," *AI Mag.*, vol. 76, pp. 57–66, 2000

[7] M. O. Riedl, "Narrative generation: Balancing plot and character," Ph.D. dissertation, Dept. Comput. Sci., North Carolina State Univ., Raleigh, NC, USA, 2004.

[8] C. Smith, Storytellers: From Mel Allen to Bob Costas: Sixty Years of Baseball Tales From the Broadcast Booth. London, U.K.: MacMillan, 1995.

[9] Mohammed Al-Maolegi, Bassam Arkok,"An Improved Apriori Algorithm For Association Rules" ,International Journal on Natural Language Computing (IJNLC) Vol. 3,

[10] S. S. Sheik, Sumit K. Aggarwal, Anindya Poddar, N. Balakrishnan, and K. Sekar, "A FAST Pattern Matching Algorithm ", *J. Chem. Inf. Comput. Sci., Vol. 44, No. 4, 2004* No.1,February2014