Interestingness Calculation for Cricket Summary Generation

Thendral. S, R. Subhashini, V. Madhan Karky

Abstract— This paper analyses the way how to make cricket summary automation more efficient. Thus, from the score card details, by adding various interestingness factors to make the summary more interesting to the user is the solution. That can be achieved by considered both the history of the teams and the current match details and analyzed by the system.

Keywords — Machine learning; cricket summary automation; artificial intelligence; Natural Language Generation, Natural Language Programing.

I. INTRODUCTION

Nowadays Cricket is the most interesting game than all other games in Asia. Thus, the huge number of people are interested to read the summary of cricket. For making automatic cricket summary, it is a must to attract the user by adding more details than the human summary writing[4,5,6]. It is achieved by adding various interesting patterns and interesting scores added to the current match score details. It contains the following two modules:

- Interesting pattern extraction
- Interesting score calculation

II. INTERESTINGNESS PATTERN EXTRACTION

Modified version of the Apriori algorithm is here used to get all the hidden patterns from the existing data. Candidate sets are produced, and pruning is provided to avoid heavy computation because of large number of candidate sets. The frequent item sets generated are used to find the association rules (patterns). The Apriori algorithm is modified to find frequent patterns in a multidimensional database. If we need to identify the interesting patterns to the match between the two countries, then the entire previous matches of that two teams are analyzed and winning rate of both the countries calculated with the following measures:

- Overall played matches
- Toss winning
- Series details
- Pitch details
- Run rate details
- Venue details
- Player details

Revised Manuscript Received on July 18, 2019.
Thendral. S, Research Scholar, School of Computing, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.(Email: thendraldots@gmail.com)
R. Subhashini, Professor, School of Computing, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India.(Email: subhaagopi@gmail.com)
V. Madhan Karky, Karky Research Foundation,Chennai, Tamilnadu, India.(Email: inbox@karky.in)

2.1.Overall played matches

The number of matches played by both the countries in cricket is considered and the winning rate of each identified. Then the number of matches between that particular two countries are taken into count and winning result obtained.

2.2.Toss winning

The number of matches won by both the countries while win the toss is considered. Then the number of matches won by two countries while choosing batting or chasing also taken into count.

2.3.Series details

The team winning rate in particular series is analyzed. The team winning rate against both countries also taken. For example, if we need to summarize world cup series final match between India and Pakistan teams, India always wins is the pattern.

2.4. Pitch details

The number of matches played by both the countries in batting and bowling pitch are analyzed and the winning pattern calculated based on that details.

2.5. Run rate details

The number of matches won by both the countries in cricket is considered and the winning run rate of each team identified to set the pattern.

2.6. Venue details

The number of winning matches by both the countries in home and away is considered to create the interestingness pattern.

2.7. Player details

The number of batsmen and number of bowlers participated in winning matches of both countries are identified to create the interestingness pattern.

Thus, with the help of the above measures, the interesting patterns obtained, and prediction done entirely based on both teams match history.

III. INTERESTINGNESS SCORE CALCULATION & RESULTS

The data to be gathered for match details is present in www.espncricinfo.com. They provide the API for entire details of the particular match. This system has a custom designed parser, for the tag structure of this site web API.

The user must provide the URL from where the particular matches data can be obtained. The system checks whether the match has already been processed. If not, the parser parses the page and retrieves the statistical data in the form of predefined feature vectors. These vectors are then populated in the corresponding tables in the database. If the particular match is already processed, the system redirects the control to the user interface which already contains feature vectors. The interestingness score is in the range from one to ten. The weightage to the feature vector is assigned by considering the following factors:

- Run rate
- Winning margin
- Interesting pattern
- Records made
- Fall of wickets
- Series state
- Score details

3.1. Run rate

The weightage given to the both teams run rate, individuals high run rate, run rate in overs, events like rapid run rate in specific crucial overs can become the turning point of the match. These events are identified and are given different weights.

3.2. Winning margin

The Facts like difference between the final scores, difference between number of wickets lost, number of balls left, difference between the number of overs played are considered while quantifying this factor.

3.3. Interesting pattern

The winning rate calculated by pattern extraction from the team's history and the prediction concluded. If the current match is differing from the existing pattern, then more weightage given to that event.

3.4. Records made

The team records and the individual player records like Half centuries, centuries, double century, quick fifties, quick centuries, hattrick, a bowler taking more than 3 wickets are given different levels of weights to contribute to the interestingness.

3.5. Fall of wickets

The number of wickets taken by both teams, took by individual players, fall of wickets frequency, wickets taken with less runs given by bowler, rapid fall of wickets in specific crucial overs can change the winning of match. Thus, these events are identified and are given different scores.

3.6. Series state

If the match is a part of a series, then the importance of the match in terms of winning or losing the series is quantified with various weights.

3.7. Score details

The number of 4s and 6s by the team and individual players are quantified here to decide the level of interestingness.

Thus, the interestingness score assigned depend on the above measures. Based on the scores the interestingness priority queue obtained. Depend on the size of the cricket summary, the number of elements from the queue will be considered. Each element is considered as an event. Each event forms a sentence in the summary text.

IV. CONCLUSION

we have created the automated cricket summary generator with interestingness measures. This analyse the entire past matches of two countries. Seven details considered to predict the winning pattern. Seven measures consider to provide score. It can further develop with adding additional details to obtain pattern and adding additional measures to obtain interestingness score. This interestingness scoring model will be used to develop summary for other games too.

V. REFERENCES

1. Alice Oh and Howard Shrobe, "Generating baseball summaries from multiple perspectives by reordering content," in Proc. the 5th International Natural Language Generation Conference, 2008, pp. 173-176.
2. Jacques Robin and Kathleen McKeown, "Empirically Designing and Evaluating a New Revision-Based Model for Summary Generation," Department of Computer Science, Columbia University, 1996, vol. 85, pp.135-179.
3. Thendral, Subhashni and Madhan Karky, "Seithiyalan – An iPhone Application for Cricket Summary Generation in Tamil", International Journal of Recent Technology and Engineering (IJRTE), Volume-7, Issue-5C, February 2019.
4. Vigneshwari, S., Mary Posonia, A., Gowri, S. " An efficient framework for document retrieval in relationship strengthened personalized ontologies ",Advances in Intelligent Systems and Computing, ISBN: 978-981-13-0513-9,pp 735-742
5. P.Ajitha, Dr G Gunasekaran,"Emotion Classification in web document using fuzzy inference system",Global Journal of pure and applied Mathematics, ISSN:0973-1768,Vol.12 No.1.2016 PP 83-93.
6. Sudha B, Jabez J., " Secured and optimal retrieval of data in cloud through computational techniques ", , ARPN Journal of Engineering and Applied Sciences, VOL. 11, NO. 13, JULY 2016