2015 International Conference on Computational Science and Computational Intelligence

Soccer Events Summarization by Using Sentiment Analysis

Said JAI-ANDALOUSSI, Imane EL MOURABIT, Nabil MADRANE, Samia BENABDELLAH CHAOUNI,
Abderrahim SEKKAKI

Abstract— **In this paper, we propose a framework for automatic textual soccer summarization based on real-time sentiment analysis of Twitter events. The interaction of fans on Twitter is marked by the expression of their sentiments through their tweets, they can be positive or negative depending on the event in soccer games. Our framework analyzes the sentiments expressed on Twitter with the goal to detect and predict the team supported by each fan as well as actors (the team(s) and player(s) concerned(s)) and the details associated with each event. All this information is used to draw up a summary of soccer matches. The summary is constructed using machine learning and KDD process that allows the extraction of knowledge from data in the context of large databases. Through the realized experiments, we confirm that the results are promising and this work has allowed us to verify the feasibility and efficiency of soccer events summarization by using sentiment analysis in media streams**

Keywords-Soccer Events Summarization; Sentiment Analysis; Text Mining; Machine learning;

I. Introduction

Nowadays, Soccer fans are posting millions of status updates to social networks at every hour of the day, such as Twitter and Facebook. As of July 2015, more than 600 million updates were being posted to Twitter per day [1]. Some of these updates represent the events that users share while watching television programmes such as: TV series, live TV shows, live breaking news and sporting events [1]. These updates provide important information about what happens in real time, but this information should be summarized and synthesized so that it's accessible. Many research works have been done in the field of twitter text mining and more particularly in the sentiment analysis. A real application was exploited for the prediction of presidential races in some countries (USA, Pakistan, ..). The generation of soccer events summary from Twitter is a relatively new area of research, but it is expanding rapidly. Several methods have been proposed: Guido van Oorschot et al [2] presented a method based on a three-step approach that allows the detection of the interesting minutes, then the type of event that occurs in these minute is classified and assigned to the concerned team. Jeffrey Nichols et al [3] have proposed an algorithm to produce a summary for sporting event using Twitter. This method is based on extraction of reasonable sentence-level summaries of important moments, and then these moment summaries can be concatenated to produce an event summary of a few paragraphs. These event summaries could communicate what happened during an event to a person who might have been unable to follow it in real time. Robert P. Schumaker applied sentiment analysis on tweets to predict the soccer match's result. They have shown that if the number of negative tweets assigned to a team is equal to 1000 this means that the team is down by one. If the number of positive tweets is between 2000 and 10000 this implies that the team made a goal. More than 10,000 positive tweets implies that the team won by 2 goals or more. The obtained results show the power of the sentiment analysis in tweets. In our previous work [4] we proposed a framework for automatic soccer video analysis and summarization by using video content analysis and social media streams. The idea is to detect the soccer events by using spikes in tweet volumes in a stream filtered by some hashtags and associate this information with result derived from video analysis. But what is missing in our old proposed approach is the textual description that will add a semantic description for this summary. This information can be extracted from the tweets collected, by using text mining and sentiment analysis. This description associated with the video summary may be used in several types of applications, e.g. the content based video retrieval ,video annotation etc... In this work we try to use sentiment analysis, text mining and other techniques to elaborate textual soccer summaries using the information extraction from tweets.

An organization chart of the proposed approach is given in Figure 1

II. Related Work

To develop the automatic summarization of events in the context of sports games, the majority of research works handles the detection of the events through the analysis of the video and audio content. In [4] the audio energy analysis of the sports commentator's speech and visual content analysis are used to extract events. This technique has been performed in several sports fields such as: soccer [5], Baseball [6] and other sports with varying success. Audio and visual content analysis are computationally expensive and generally an event can be extracted, but not classified. There are approaches that couple the video signal with a textual information such as a minute-by-minute report [7], but these reports still need human input. Lately, crowdsourced data have been proposed to deal with this problem. [8] present a mobile application in which fans can annotate events in a soccer game. The results proved that the number of annotations has been increased around many significant events in

Said JAI-ANDALOUSSI, Imane EL MOURABIT, Nabil MADRANE, Samia BENABDELLAH CHAOUNI and Abderrahim SEKKAKI are with Faculty of sciences, Hassan II University - Casablanca, Morroco. email : said.jaiandaloussi@etude.univcasa.ma

[1] http://www.tweetstats.com

978-1-4673-9795-7/15 $31.00 © 2015 IEEE
DOI 10.1109/CSCI.2015.59

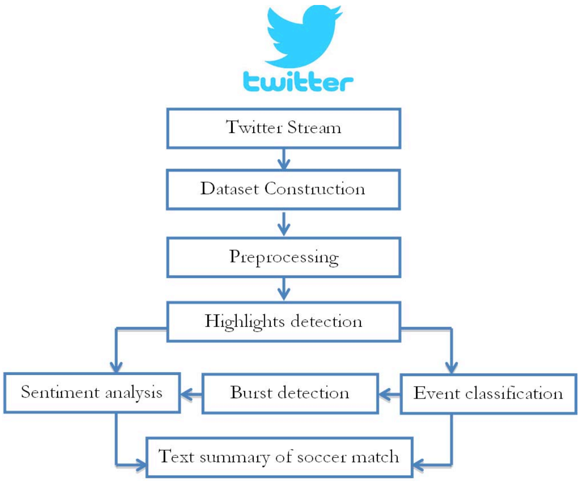


Fig. 1. Organization chart of soccer text summarization

a game, but people still needed to make a particular effort to use the application. On Twitter, fans are discussing soccer games with each other. In our old [4] work we used the data extracted from twitter for events detection in soccer matches. We employed a simple approach detecting significant minutes by finding the peaks in the stream of Twitter. Our results are comparable to events detection from audio and video, but still suffer from some false positives. In this work we aim to improve the quality of the results by using burst detection based on moving-threshold and some algorithms of machine learning to classify the events and enriching the event information by assigning it to the concerned team. Doing this, we seek to develop an automatic summarization system which allows to extract relevant information from a large number of tweets using sentiment analysis, without the need of conventional solutions that require a strong human involvement. For that, we have adopted the KDD (Knowledge Discovery in Databases) process to develop our system. The term Knowledge Discovery in Databases [2], or KDD for short, refers to the broad process of finding knowledge in data, and emphasizes the "high-level" application of particular data mining methods. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems, and data visualization. The unifying goal of the KDD process is to extract knowledge from data in the context of large databases. It does this by using data mining methods (algorithms) to extract (identify) what is deemed knowledge, according to the specifications of measures and thresholds, using a database along with any required preprocessing, subsampling, and transformations of that database.

III. Data

In this section, we will present the data collection process as well as the data preprocessing steps.

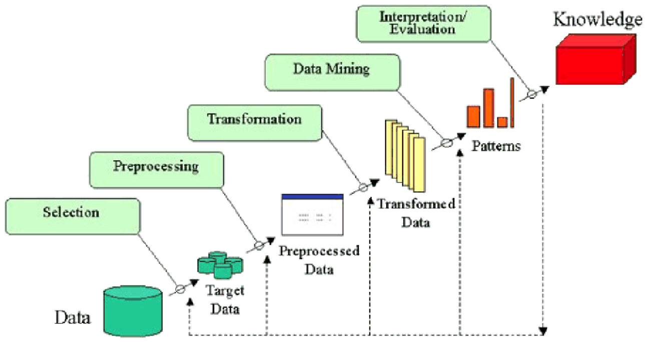


Fig. 2. An Outline of the steps of the KDD process

A. Data collection

To collect data, Twitter offers several APIs for querying its database. In this work we chose to use Streaming API implemented by twitter4j[3], this API returns only the data created in real time, i.e. the tweets created just after sending the request and which perfectly meet our needs. Contrary to [2] that has chosen to collect data 15 minutes before the start of the match and after they seek information "has started "to detect the start of the game, we decided to send the request just after the kick-start of the match and stop the request for half-time. Experience shows that people tweet a lot before starting the match to express their expectations of the matches, the final prediction score and during halftime they speak of past events in the match which were already commented, that pushes the tweets number to reaching a certain threshold and the system will detect the events that are false positives.

To collect data from Twitter several works have proved that the use of hashtags is the most effective means to gather tweets around a specific topic. In our work, we have seen that by convention hashtags are created that consist of abbreviations of club names for each soccer game, beginning with the home team. For that, We extract hashtags from the official Twitter accounts of each team, these hashtags follow the convention: $< HomeTeam > v < AwayTeam >$ e.g #HCFCvAFC, #AFCvSCFC, #MUFCvAFC, #AFCvSFC, #AFCvWBA.
In this work, we use three types of hashtags:

- $\# < HomeTeam > v < AwayTeam >$
- $\# < HomeTeam >$
- $\# < AwayTeam >$

As part of this work, we are interested in England premier league, this championship is one of the most prestigious in the world and the most popular in terms of viewers and fans, estimated at over one billion in 2007 and more all teams of this championship are present on twitter since 2008, they have official accounts with millions of subscribers.

[2] http://www2.cs.uregina.ca/~dbd/cs831/notes/kdd/1_kdd.html

[3] http://www.twitter4j.org

B. Data cleaning and preprocessing

This step is very important in order to prepare the data so that they are usable in the rest of the process.

1) Spam suppression: Social media are increasingly being targeted by spammers. According to several studies, 40% of accounts created on Facebook, Twitter and others are used for spam, this is called social spam. In this step of data preprocessing we chose to classify our data into spam and not spam using a supervised learning method and for training data, we use the most popular spam on Twitter, then our program will detect and delete spams. Supervised learning algorithms are performed on 5567 entries, each entry is assigned to a spam or not spam label, we used three learning algorithms: SVM (Support Vector Machine), Naive Bayes and neural network proposed by the WEKA Framework [4] with 10 folds of cross-validation. To evaluate our system, we manually labeled 1000 significant tweets from several matches of premier league and we compared these algorithms. The experiments showed that the Naive Bayes algorithm gave better performance for the classification of spam, consequently we adopted this algorithm for the spam detection.

2) Detection of inactive users and media Twitter accounts: Each Twitter API provides us information about the person who tweeted, we decided to use the information of "username" to compute the tweets, number of each user in a soccer match to determine whether it is active or not and decide to take into consideration his tweets or reject them. We choose to reject all users who are under 4 tweets in the database because they are inactive users. The experiments showed that users who have more than 50 tweets in the database represent television/ radio program accounts, or sports websites, their tweets will also be rejected as we seek to analyze the tweets generated by simple viewers. For that, we looked for users who who have between 4 and 50 tweets in the database.

3) removing irrelevant data: Generally, all tweets contain metadata that can determine whether the tweet is the retweet of another tweet or not. In our case, the retweets will be rejected because they do not provide new information, It will be exactly the same for the tweets containing urls and videos. To homogenize the information collected from the tweets, we chose to keep only the tweets that are written in English.

4) text normalization: in order to normalize the text of tweets, we apply the following rules for cleaning all the collected tweets:

- convert all words in lowercase
- delete the repetition of words
- replace slang words by their equivalents using a dictionary[5] e.g LoL → Laughing out loud;
- the lemmatization used to group words from the same family (am, are, is) → be; (car, cars, car's) → car;
- Delete the stop words e.g (so, Some, then, the, a, about, no), emoticons, hashtags and punctuation.

original tweet	tweet after preprocessing
GOOOOOOAAAAALLLL! ! BIG MAN NASRIII ¡3 :D WHAT A GOALLL ! 1-0 ! !	goal big man nasri goal 1-0
Loool ! ! ! Ward does brilliantly to win Palace a corner which Puncheon will take. #CPFC	laughing out loud ward do briliant, win palace,corner puncheon will take
Yellow caaaaaaaaaaaarrrddddd ! ! ! ! ! ! ! Damien Delaney is shown the first yellow card of the match, for a poor step-in on Sergio Aguero. #CPFC #MCFC	yellow card damien delaney be show first yellow card match poor step-in sergio aguero

TABLE I
EXAMPLES OF TWEETS BEFORE AND AFTER THE PREPROCESSING

we've carried out these treatments with WEKA Framework. To enrich the data we used the NLP Stanford (Natural Language Processing) [6] and the enrichment is performed by the part-of-speech tagging.

C. Data Transformation

In this step we are looking for methods of dimension reduction and transformation to select the relevant words, thereby reducing the dimension of the vector space and increases the efficiency, indeed the dimension of the vector space acts on the complexity of the learning model. The more words we take into account, the more variables will have to be estimated and thus, the further the probability of error increases. For these reasons, We performed to reduce the vector space, taking into account the preprocessing made in the previous step and we choose only the adjectives, adverbs, verbs and names generated by the part-of-speech tagging like features of the tweet text, because it is only those features that have important roles in sentences.

In order to extract significant words for each tweet, we apply tf-idf method weighting approach to rank the words. Tf-idf, short for term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is often used as a weighting factor in information retrieval and text mining. The tf-idf value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

The tf and idf are simply expressed by the equations below :

$$tf(t) = \frac{Nb_t}{\sum Nb_t};\ \ idf(t) = log(\frac{Tn_d}{Nb_d}) \tag{1}$$

$$tfidf(t) = tf(t) * idf(t) \tag{2}$$

where
Nb_t: Number of times term appears in a document
$\sum Nb_t$: Total number of terms in the document
Tn_d: Total number of documents

[4] http://www.cs.waikato.ac.nz/ml/weka/

[5] www.internetslang.coml

[6] http://nlp.stanford.edu/

Nb_d: Number of documents with term t in it

IV. Highlights Detection

A. Events classification

Our approach is based on a classification of tweets into five classes (goals, red card, yellow card, penalty, foul), we have opted for using supervised learning algorithms, so the first task was to label tweets that we will use as a training set.

As we sought to classify tweets for several soccer games, we found that labelling a dataset of each game was heavy, complex and not a practical task. Therefore we chose to construct a common training dataset and we collected tweets from several soccer games, then we chose tweets that describe the class to which they belong. Next, we deleted from the tweets text team names, names of players, hashtags and any other information that could identify the game from where the tweet is extracted. At the end, we manually assign each tweet to a class.

As in the classification of spam, we have used the implementations of algorithms SVM, Naive Bayes and Neural Network proposed by the WEKA Framework with cross validation 10 folds. To evaluate our classification program, we manually labeled 1000 significant tweets and we computed the measures presented in table III.

F-measure is the harmonic mean of recall and precision, where recall is the ratio of polar examples that are found by a method, and precision is the percentage of correctly found polar examples amongst the ones that a polarity detection system marked as polar: Those two cases illustrate that the SVM algorithm gives the best performance for the classification of goals, we generalize the result and we adopt this algorithm for the detection of other events.

$$F-measure = 2 * \frac{precision * recall}{precision + recall} \tag{3}$$

B. Burst detection based on moving-threshold

In this paper, we use a moving-threshold burst detection technique [4] that allows the detection of important events in a soccer game. This technique uses a moving-threshold that is obtained by computing the standard deviation and mean of the number of tweets during the playing of soccer games. For the implementation of this approach, we define a sliding window with the length l. We find that in our experiments the length l equal to 15 seconds gives the best results. In the time sequence $(t_1, ...t_n)$ during the playing of soccer games, we have a sequence of sliding windows $(l_1, ...l_n)$. The sliding window l_i at time t_i contains $N(l_i)$ tweets. The algorithm is given below:

1) the length of the sliding window is l
2) At the time t_i, obtain the sliding windows $(l_1, ...l_i)$ and the corresponding numbers of tweets $N(l_1), ...N(l_i)$, and compute the values:
   - $mean_i = mean(N(l_1), ...N(l_i))$
   - $std_i = std(N(l_1), ...N(l_i))$
   - $MT_i = \alpha * (mean_i + x * std_i)$
3) The highlight in soccer games is detected in $l_i \;\; if \;\; N(l_i) > MT_i$

where α is the parameter used to relax the condition of moving-threshold and x is a constant defined between 1.5 and 2.0.

C. Sentiment analysis

Twitter is a good data source for sentiment analysis because the tweet size is limited to 140 characters and contains different sentiments, especially during soccer matches fans express their sentiments for the best events (goals, red card, yellow card, penalty, foul).

The objective of sentiment analysis is to analyze large amounts of data in order to deduce the different sentiments expressed therein. Generally, there are three different methods of automatic sentiment analysis. The first method is based on a lexicon built from existing dictionaries (lexicon-based approach).The second method is based on a corpus containing evaluative texts whose language is generally subjective (corpus based approach). The third method is based on evaluative texts, it treats sentences semantically by applying a deep learning approach. We have tested these three approaches:

- method 1 was implemented with the Framework R[7]
- method 2 has been implemented with the Framework weka
- method 3 has been implemented using the Stanford NLP

The table II shows the examples of obtaining results for the three approaches. We note that the third method detects the sentiments with degrees (0: Very negative; 1: Negative; 2: Neutral; 3: Positive; 4: Very positive). Several released experiments show that the third method gives the best performance, we've thus adopted it.

Football is one of the best cases to analyze sentiments due to the symmetric nature of events what is good for one time is bad for the other. Usually fans express what they feel about soccer games events and react in real time, positive feelings are always generated after the highlight which made them happy like goals and negative feelings are generated after the highlight which made them unhappy like red cards. Our approach consists to identify and understand these feelings that can provide valuable results and predictions, for this reason we apply the burst detection algorithm to detect minute/highlights per class next we selected tweets created in these minutes to analyze their sentiments then we detect named entities and we assign an event to a team (following section).

V. Text summarization

In this step, we use the moving threshold burst detection method to detect the minutes of important events in a soccer game [4]. After that, We try to affect each event to the concerned team.

[7] https://www.r-project.org

Tweets	method 1	method 2	method 3
A horror back pass from Phil Jones who almost kicks it into his own goal. De Gea to the rescue. Nervy start	Neg	Pos	Neg=1
I,was excited to watch this game but it is not a good start	Pos	Pos	Neg=1
Absolute horror show , this is really awful. wake up spurs,this is the worst performance I have seen	Neg	Neg	Neg=0
Great performance, great Goals by Fellaini, Carrick and Rooney, our players were brilliant.Proud. amazing goal amazing celebration	Pos	Pos	Pos=4
Shameful,embarrassing, abysmal, disappointing, desireless !	Neg	Pos	Neg=0
Clumsy phil jones ! It?s shameful as a fan to watch this	Neg	Neg	Neg=1
great result, brilliant goals an incredible celebration	Pos	Neg	Pos=4
I wonder who is more nervous right now : me or ? I?m guessing we?re equally terrified of this game	Neg	Neg	Neg=3
The,first half was terrible and Horrible but the second half made me happy	Neg	Neg	Pos=3
its a great save from Fellaini but I didnt find it spectacular	Pos	Pos	Pos=2
Tweets correctly classified	**7/10**	**4/10**	**10/10**

TABLE II

SENTIMENTS DETECTION USING THREE DIFFERENT APPROACHES

A. *Named entity detection*

The named entities are the types of tokens representing entities of the concrete world, in the context of soccer, examples of named entities are the names of the players (e.g Toure, Fellaini, Nasri, etc...) names of teams (e.g. Manchester United, Chelsea, etc...), the names of the football stadium (e.g Emirates Stadium, Wembley Stadium, etc...). We chose to detect the names of players by the Named Entity tagger of Stanford, which is based on a machine learning. The training data are collected from texts containing proper names. We also tested the NameFinder OpenNLP of Apache, the main drawback of these detection tools named entities that they are based on data containing proper nouns and the presence of the first capitalized letter. To solve this problem, we created our own proper nouns detector that is not based on a machine learning, but it is based on a dictionary that we collected, it contains, up to, now 415 players names with corresponding teams

B. *Team Assignment*

In the previous experiment, we were able to detect with fair accuracy events having taken place in a certain minute of soccer game, but it's yet unknown which team scored the goal or received the red card etc. Soccer is a symmetric game, where fans of a team are likely to be happy if their team scores, fans of the adversary team are expected to be unhappy when faced with the same event. For this reason, we analyzed sentiments in our dataset using Stanfords NLP. To assign an event to a team, we selected events minutes and for each minute we selected tweets whose opinion are positive to detect if there's a goal and tweets whose opinion is negative to detect other events, then we detected user's favorite team, we admitted that if a user tweets positively using slangs team such as #KTBFFH #COYB, #COYI, #COYG he's an immediate fan of this team as we compute the number of times when he tweets positively using the official hashtags of the two teams (e.g. #MUFC, #HCFC #AFC..). The higher frequency will prevail, then we divide our dataset into two datasets: tweets intended to the first team and tweets intended to second team then we compare the length of datasets if the length of the first team's tweets is greater than tweets of second team's we assign the event to the first team and vice versa.

C. *Summary generation*

After the selection pre-processing, classification and sentiment analysis. we need to use all the information coming from this process to generate a text summary that describes the highlights of a soccer game. There are several methods that have been used to generate a summary text from tweets, save that the latter might contain spelling errors, letter repetition, information incomplete and swear words expressed by some fans. Jeffrey Nichols et al [3] are generated summaries from the tweets, but the drawback of their system is the lack of information to present. For the generation of summary We find that the final score did not reflect the reality of the events that occurred during a soccer game. On these grounds, we have opted to use sentence templates to build our summary. For that, our system selects the words and sentences depending on: 1) final match score, 2) number of received tweets in each half-time, 3) type of highlights and other information, in order to prepare an extensive summary that expresses exactly the conduct of the game in terms of the highlights.

VI. RESULTS AND DISCUSSION

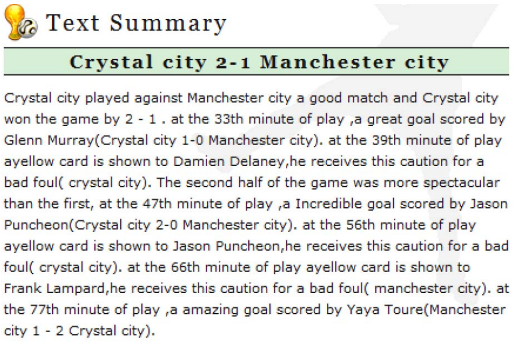
In this work, we use three methods to evaluate our system:

1) we compare our summaries to gold standards generated by humans
2) we check the description generated for the events and their importance, in other words, the goals and how they are scored (spectacular, amazing), the seriousness of players' injuries, etc...;
3) our manual evaluation of the summaries for grammatical correctness, readability and meaning in comparison to the gold standards.

We use two gold standards:

- Game recap articles written by BBC Sport, www.francefootball.fr and www.matchendirect.fr
- Manual summaries generated by ourselves for each important moment detected by our system.

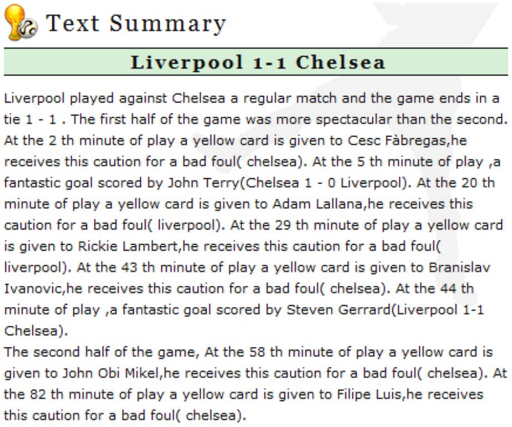
The recap articles allow us to evaluate our summary of the soccer game, although these articles are longer than the summaries generated by our system. The manual summaries generated by humans allow us to evaluate our summaries of each important event separately. we give in figure 4 an summarie example generated by our system. Table IV shows the detection accuracy percentage of the most important events (goals, red card, yellow card, penalty, free-kick) in a soccer matches, we clarify that the results are obtained around 100 matches of the england premier league. As you can see the the mean precision of soccer video summarization, the percentage reaches 90,04%.


Text Summary

Crystal city 2-1 Manchester city

Crystal city played against Manchester city a good match and Crystal city won the game by 2 - 1 . at the 33th minute of play ,a great goal scored by Glenn Murray(Crystal city 1-0 Manchester city). at the 39th minute of play ayellow card is shown to Damien Delaney,he receives this caution for a bad foul(crystal city). The second half of the game was more spectacular than the first, at the 47th minute of play ,a Incredible goal scored by Jason Puncheon(Crystal city 2-0 Manchester city). at the 56th minute of play ayellow card is shown to Jason Puncheon,he receives this caution for a bad foul(crystal city). at the 66th minute of play ayellow card is shown to Frank Lampard,he receives this caution for a bad foul(manchester city). at the 77th minute of play ,a amazing goal scored by Yaya Toure(Manchester city 1 - 2 Crystal city).

Fig. 3. summary example of the match of Crystal city - Manchester city


Text Summary

Liverpool 1-1 Chelsea

Liverpool played against Chelsea a regular match and the game ends in a tie 1 - 1 . The first half of the game was more spectacular than the second. At the 2 th minute of play a yellow card is given to Cesc Fàbregas,he receives this caution for a bad foul(chelsea). At the 5 th minute of play ,a fantastic goal scored by John Terry(Chelsea 1 - 0 Liverpool). At the 20 th minute of play a yellow card is given to Adam Lallana,he receives this caution for a bad foul(liverpool). At the 29 th minute of play a yellow card is given to Rickie Lambert,he receives this caution for a bad foul(liverpool). At the 43 th minute of play a yellow card is given to Branislav Ivanovic,he receives this caution for a bad foul(chelsea). At the 44 th minute of play ,a fantastic goal scored by Steven Gerrard(Liverpool 1-1 Chelsea).
The second half of the game, At the 58 th minute of play a yellow card is given to John Obi Mikel,he receives this caution for a bad foul(chelsea). At the 82 th minute of play a yellow card is given to Filipe Luis,he receives this caution for a bad foul(chelsea).

Fig. 4. summary example of the match of Liverpool - Chelsea

Chelsea ≠ Sunderland 24 Mai 2015	SVM	Naive Bayes Bayes	NN
Precision(%)	94.73	80	76.34
Rappel(%)	96.77	92.30	95.94
F-measure(%)	95.73	85.71	85.02
Accuracy(%)	92	80	75

Liverpool ≠ Chelsea 10 Mai 2015	SVM	Naive Bayes	NN
Precision(%)	96.59	72.22	69.44
Rappel(%)	97.70	86.66	75.75
F-measure(%)	97.14	78.78	72.45
Accuracy(%)	95	68	62

TABLE III

EXAMPLES OF EVALUATION MEASURES FOR THE GOALS CLASS

	Goals	Red card	Yellow card	Penalty	Fool
Precision(%)	100%	100%	86%	100%	66%

TABLE IV

SYSTEM EVALUATION MEASURE

CONCLUSION

In this paper, a framework for summarization of soccer match has been presented. This framework allows soccer highlight detection and summarization by using sentiments analysis and text mining (Twitter). Several methods have been proposed in this work, such as, the KDD approach, machine learning algorithms, the moving threshold burst detection algorithm used to detect the bursts of tweets on Twitter and other techniques. The testing results proved that, the proposed framework is efficient with a mean precision of 90,04%. From this result, we believe that the proposed approach can play an important role in improving the quality of the summarization of soccer events and the proposed framework can be applied to other sports, such as baseball, handball and basketball.

REFERENCES

[1] WOHN, D. Yvette et NA, Eun-Kyung. Tweeting about TV: Sharing television viewing experiences via social media message streams. First Monday, 2011, vol. 16, no 3.

[2] VAN OORSCHOT, Guido, VAN ERP, Marieke, et DIJKSHOORN, Chris. Automatic extraction of soccer game events from twitter. In : Proc. of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web. 2012. p. 15.

[3] NICHOLS, Jeffrey, MAHMUD, Jalal, et DREWS, Clemens. Summarizing sporting events using twitter. In : Proceedings of the 2012 ACM international conference on Intelligent User Interfaces. ACM, 2012. p. 189-198.

[4] JAI-ANDALOUSSI, Said, MOHAMED, Aboulaite, MADRANE, Nabil, et al. Soccer Video Summarization Using Video Content Analysis and Social Media Streams. In : Proceedings of the 2014 IEEE/ACM International Symposium on Big Data Computing. IEEE Computer Society, 2014. p. 1-7.

[5] Li, B., Pan, H., Sezan, I. (2003, April). A general framework for sports video summarization with its application to soccer. In Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on (Vol. 3, pp. III-169). IEEE.

[6] Li, B, Sezan, M. I. (2001). Event detection and summarization in sports video. In Content-Based Access of Image and Video Libraries, 2001.(CBAIVL 2001). IEEE Workshop on (pp. 132-138). IEEE.

[7] Xu, C., Wang, J., Wan, K., Li, Y., Duan, L. (2006, October). Live sports event detection based on broadcast video and web-casting text. In Proceedings of the 14th annual ACM international conference on Multimedia (pp. 221-230). ACM.

[8] Sahami Shirazi, A., Rohs, M., Schleicher, R., Kratz, S., Mller, A., Schmidt, A. (2011, May). Real-time nonverbal opinion sharing through mobile phones during sports events. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 307-310). ACM.