Noname manuscript No.
(will be inserted by the editor)

On Semantic Annotation for Sport Video Highlights by Mining User Comments from Live Broadcast Social Network

Po-Fang Hsu[1] · Yao-Chung Fan[1] · Huan Chen[1]

Received: date / Accepted: date

Abstract In recent years, the idea of viewing online social media as human-powered sensing networks has draw significant attentions in research communities. Great examples are Twitter-based earthquake detection, Influenza detection, and traffic abnormally detection. Following the same viewpoint of the human-powered sensing network, in this paper, we discover the utility of user-generated social texts on social media platform for extracting highlights and annotating the semantics of sport video clips. The basic idea for the leverage of social text is that one can make use of the semantics of the social texts for understanding the corresponding moments of the game. For example, when watching a baseball, the users on social media will timely comments about the play, the team, and the events. By properly analyzing the texts, automatically annotating the sport videos turns out to be possible. However, two research challenges need to be addressed for such an idea: 1) as sport videos are often lengthy, how to precisely locate the moment of important events is a challenge task, 2) social media contents are generated by users on social network platform and contains various information and with noises, and therefore how to distill useful information from noisy social comment is also a challenge. In this paper, we present a weighting scheme to address the issues by estimating the importance of users (and therefore their comments) on social network platforms based on mining the interaction between users on social platforms. Also, we use soccer game videos and baseball game videos as well as social comment from on-line social network as our test data set. The evaluation over real data shows the effectiveness of the proposed framework.

Keywords Social Media · Semantic Annotation · Sport Video Summarization · Personalized Information Retrevial.

Po-Fang Hsu
E-mail: daniel_hsu@smail.nchu.edu.tw

Yao-Chung Fan
E-mail: yfan@nchu.edu.tw

Huan Chen (✉)
E-mail: huan@nchu.edu.tw

[1]Department of Computer Science and Engineering,
National Chung Hsing University, No.250, Guoguang Rd.,
South Dist., Taichung 402, Taiwan

1 Introduction

In recent years, the applications of viewing online social media as human-powered sensor networks has draw significant attentions in research communities. The viewpoint is that social media can be a source providing timely information about real world events of all kinds by mining the data produced by users' activities on the social media platform. Great examples are Twitter-based earthquake detection [1], influenza detection [2, 3], and traffic abnormalities [4].

With the idea of leveraging the human-powered sensor network, this paper investigates the utility of user-generated social texts on social media for annotating the semantics of important moments of broadcast sport videos. More specifically, the problem can be stated as follows. We are given (i) a sequence of user-generated texts, e.g., comments or hashtags, on a social platform about some sport game, and (2) a broadcast video of the corresponding sport game. The goal is to develop an automatic mechanism to extract highlights and to annotate semantics on clips in a broadcast sport video.

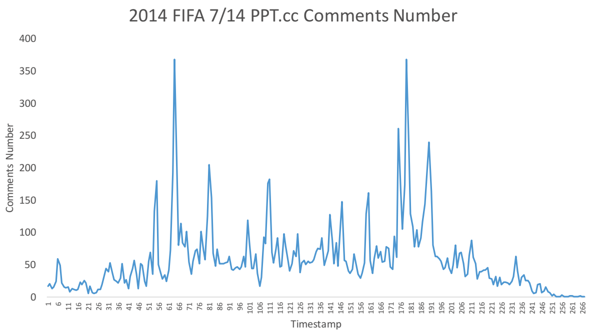

Fig. 1 The number of comments over time

Such problem is fundamental to sport video summarization and personalized information retrieval, as it would be attractive for providing personalized highlight based on users' preferences.

The insights for addressing automatic semantic annotation is that one can make use of the semantics of the social texts for realizing what happened in corresponding moments of the game. For example, when watching a live baseball game, the users on social media will timely comment about the play, the team, and the events. By looking at the words that users are using within their comments, creating annotation labels that effectively describe the event occurred at the moment is therefore feasible.

In addition, by looking at the number of comments produced in a time slot, devising a method for automatically extracting a highlight is also possible, as when an interesting event take places, users might acclaim and discuss/comment the play. For example, in Figure 1, we show the volume of social comments of a game over time, where y axis is the number of comments post, and x axis the time from the start to the end of the game. One can observe that there are fluctuation over the game. The main idea is that one can treat the number of comments over time as a time series, and the variation of the time series, such as a significant number of comments emerged, may reveal events in a game, which therefore can be employed to identify the interesting moments in the game. Techniques for sport video summarization based on such ideas have also been proposed by several pioneering studies [5, 6].

While the idea looks promising, the existing techniques have two deficits. First, the data on social network platforms are with noises; social media contents are formed by users having different comment behaviors; some are eagerly posted, and some are posted only at important moments. Bias will occur if we simply count the numbers of comments for detecting highlights. However, the issue of how to clean the noisy data was not addressed by the existing techniques. Second, there are numerous comments made from the crowd on social network platforms. The issue of how to extract useful and effective comments from a huge number of comments on social platform is also ignored by the existing techniques. For example, from the experimentation, we find that there are nearly 200 comments leaved at one time slot. Thus, a substantial information filtering effort is required to successfully drill down the social comments to relevant topics and events .

In this paper, we present a weighting scheme to address the above-mentioned issues by estimating the importance of users (and therefore their comments) on social network platforms. There are two benefits for doing so. First, by weighting users, the number of candidate comments for annotation can be significantly reduced by considering only the comments from experts (i.e., users with high weights). Second, with the importance weighting, effective event detection can be performed by considering the moments that many experts leave comments rather than the moments with a large number of comments (as proposed by the existing studies). The basic idea for weighting users is that an important user is the one who comments and participates the discussion in an important moment, and an important moment is the one with many important users leave their comments. Based on such an idea, we propose an iterative algorithm to estimate the weights (as importance measurements) of users and moments.

The contributions of the paper are summarized as follows:

– Our weighting scheme is independent to language type and sport type by leveraging the relationship between users and event moments. Our scheme differs from the existing schemes, e.g., [6] by without by requiring the analysis of social text contents, such as sentiment analysis.
– A novel recursive mutual enforcement algorithm is proposed based on a bipartite model for learning users' importance of a sport game and adaptively detecting highlight boundary.
– We use soccer video and baseball video as our testbed. The evaluation over real data shows the effectiveness of the proposed framework. Our approach is generic and can be readily extended to other sports.

2 Related Work

In this section we review the existing method of sport video summarization and semantic annotation.

2.1 Event Moment Detection

In recent years, extensive research efforts have been made to sports video summarization. The existing approaches can be divided into two categories: analyzing by video content only and analyzing with external knowledge. In the following, we review the existing works along the two directions.

Most of previous works on sports video summarization are based on employing low-level features, such as audio or visual features, extracted from video content itself. The basic idea is to extract the low level features and then performs rule-based or machine learning algorithms to detect events in sports video. For example, in [7] authors employ audio features for baseball and soccer event detection. The study of using visual features for soccer event detection is addressed in [8]. Textual features were utilized in [9] for baseball and soccer games. However, the content-based sport video summarization all suffer from the shortcoming of heavily relying on audio/visual/textual features extracted from the video itself. The semantic gap between low-level features and high-level events are the main concern for the accuracy of the event detections.

As mentioned, content-based sport video summarization faces the computational difficulty in processing video content using computer vision and image processing techniques. Therefore, recent researches have proposed techniques by leveraging the external knowledge, such as webcast text [10] or closed caption [11] to sport video summarization. For the research based on closed caption, the main idea is to make use of caption text on the video to perform sports video summarization. However, the approaches based on caption text suffer from two concerns. First, the accuracy of recognizing caption texts is affected by video quality. Second, the caption text is a simply transcript from speech to text, which may be indirectly relevant to the video content. The idea of the work using webcast is to make use of webcast text, which is a short text reporting the status for sports game and easily obtained from the web, and therefore without the problem of recognizing caption text. However, the research based on the webcast text still require low level features to detect the boundary of the events. Differing from the existing research, we start from the angle of employing social media text stream, e.g. opinion comment posts to detect semantic events in live sport videos to the summarization problem. The idea of leveraging the power of social media has also been proposed by [12–14]. In [12, 14], the idea of using bursts in twitter streams to detect events is proposed at the same time. And, in [13], the authors propose an SVM feature analysis technique to detect shot boundary. While the ideas are neat, the works all assume the data from social media is without noises, which is not true in practice. The study in [6] also proposes to weight users and moments, as we proposed. However, the proposed method requires content analysis over social texts, which brings the language-dependent concerns.

2.2 Sport Video Semantic Annotations

In addition to the highlight extraction, the semantic annotation on extracted highlight is also a critical issue for highlight indexing and retrieval. Existing research on sports video annotation can be categorized into three

Related work on sports video annotation can be classified into structure based annotation, event based annotation and ontology based annotation. Zengkai Wang et al. developed a visualize and sound analyzing system to recognize specific event view shot [15] and Changsheng Xu et al. presented a system for highlight detecting using digital clock overlaid on the video as time stamp to map official website game record [16]. Since sports video follows certain production rules when it is broadcasted, the structure information (i.e., frames/shots) and the semantic information (i.e., plays/breaks) obtained from the structure can be used to annotate sports video. Various approaches have been proposed for sports video structure analysis both at frame level and shot level. Ahmet Ekin et al. got shots and classified them into four types according to some low-level soccer video visual feature then applied the dominant related rules and game video cinematic techniques to discern among three kinds of events selected by humans [17]. J.Assfalg et al. employed Hidden Markov Models trained by identification the part of the playing field currently framed and extraction of camera motion to discriminate between three types of highlight [18]. Recently, some researches have embarked in wisdom of the human for the feature extraction. In [19], Fabio Sulser et al. developed a system using crowdsourcing to integrate all of the annotation entered by crow workers for event annotation. Although this way gain a rather good performance, it takes significant cost. However human wisdom can be extracted from other way such as the comments in the social network.

3 The Proposed Algorithm

In this section, we present our algorithm for weighting users based on their relationship between the moments of leaving comments on social networks. In Subsection 3.1, we introduce the data model and terminology we

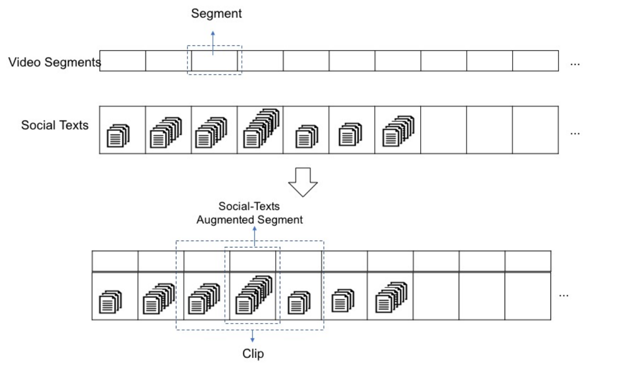

Fig. 2 Information Units: There are three information units in our data model: (1) Segment, (2) Clip, and (3) Augmented Segment

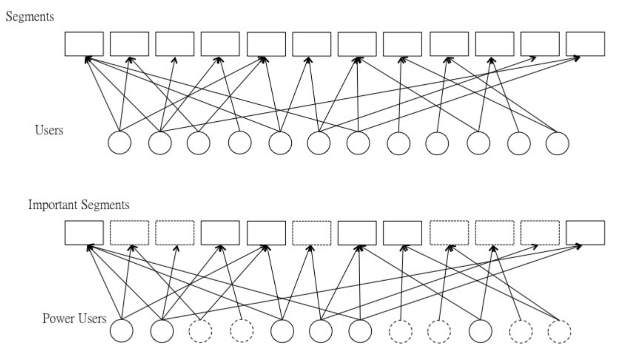

Fig. 3 A Bipartite Formulation for Users and Clips Modeling

use in this paper and in Subsection 3.2 we formulate the problem targeted in this paper.

3.1 Data Model

The data model considered in this paper is as follows. We are given two data sources: (1) a time-ordered sequence of texts (called social comments), each text message consists of two records, the user id u_i and the text contents, and (2) a broadcast sport video, which is a time-ordered sequence of video segments $s_0, s_1, ..., s_m$. We assume two data sources are time synchronized. Furthermore, by aligning the video segments with the text sequence, we augment a video segment by associating the social comments leaved in its timeline; we refer to such a video segment with social comments as *augmented segment*, which serves as basic building block in this study.

3.2 System Goal

With a given sequence of augmented segments, the goals in this paper are two folds. The first one is to discover a subsequence of video segments, within which an interesting event/play occurs. In the following discussion, we refer to a subsequence of video segments as *a clip*, and a clip contain significant plays or interesting events as *a highlight*. The relationship between the information unit is shown in Figure 2. Second, for an extracted highlight, our goal is to further annotate the highlight with a semantic description providing the following information: (1) the players involved in the highlight and (2) the terms related to events occurred. For example, a highlight might be annotated with (1) Suzuki Ichiro and (2) Home Run.

The intuition for highlight detection and annotation over the data model is that (1) by tracing the number of comments of augmented segments over time, the moment that an interesting event can be discovered, and (2) by leveraging the comments in a clip, automatically annotating the clip with play and event information is therefore possible. In the next subsection, we present a naive algorithm based on such intuition.

3.3 Threshold-based Scheme

For a given sequence of augmented segments, one idea for highlight extraction is to first compute the number of texts associated with each augmented segment, and select the augmented segment s_i whose text number exceeding a predefined threshold α as a seed. For each seed, we merge $s_{i-\alpha}, ..., s_{i-1}, s_i, s_{i+1}, ..., s_{i+\alpha}$ as a clip C_j, where α is a given parameter defining the length of a clip. If two clips are overlapped, we merge them as a new one. A naive idea is to report the extracted clips as highlights, as these clips are with much more comments than others, and therefore it is reasonable to assume that events occur in the clips.

However, such a naive approach is with two concerns. First, determining a proper threshold is a challenge task. In addition, the threshold should be also adaptive over time rather than a fixed one, as there are different number of users over the time span of a sport game. Second, our annotation process for a highlight is also based on the comments made in the highlight. The clip extracted by the threshold-based scheme contains a large number of comments, which complicates the annotation process of selecting terms for effectively describing the event happened.

For the first concern, one can employ the following equation for adaptive threshold computation, as proposed in [12]. The main idea is to adaptively determine a threshold based on a sliding window of size w over the sequence of augmented segments. Specifically, given a sequence of augmented segments $(s_1, ..., s_n)$ and a sliding window parameter w, an adaptive threshold ϕ_i at a time slot i is defined over a set s_{i-w}, s_{i-w+1}, ..., s_i by the following equation, where μ and σ is the mean and standard deviation of $\{s_{i-w}, s_{i-w+1}, ..., s_i\}$ and β is a given parameter.

$$\phi_i = \mu + \beta * \sigma \tag{1}$$

The basic intuition is to report the time slot at which the number of comments is with a significant deviation from the average value over a sliding window. Algorithm 1 shows the details about the overall process.

Algorithm 1 Threshold-based Highlight Extraction Algorithm

Input: A sequence of augmented segment s_i, i = 1, ..., N
Output: A set of clips
  for i=1, ..., N **do**
    $\phi_i := \mu + \beta * \sigma$
    if $|s_i| \geq \phi_i$ **then**
      merge $s_{i-\alpha}, ..., s_{i-1}, s_i, s_{i+1}, ..., s_{i+\alpha}$ as a clip and insert into candidate set S
    end if
  end for
  if $C_j \cap C_{j+1} \neq \emptyset$ **then**
    Extract C_j and C_{j+1} from S
    Merge C_j and C_{j+1} as a new clip and insert the clip into S
  end if
  return S

The moving threshold scheme may address the concerns of determining a threshold for detecting highlights. However, for the concerns of distilling useful information from a large number of comments in an extracted highlight, the moving threshold scheme fails to handle it. By experimentation, we found there are about 230 comments in average for highlight extracted by the moving threshold scheme.

3.4 Mutual Enforcement Weighting Scheme

In this subsection, we introduce a novel weighting scheme that estimates the importance of users as well as the importance of the segments, and then employ the users and the segments with high weights as seeds for highlight detection and highlight semantic annotation. The idea is to weight users and segments by analyzing their inter-relationship. More specifically, we can formulate the relationship between users and segments by a bipartite graph $G(V, E)$, where nodes V can be classified into two groups; one side represents users, and the other stands for segments of a game, as illustrated in Figure 3. A direct link $e \in E$ from a user to a segment represents that the user made comments about the segment.

On the basis of the bipartite formulation, we are able to weight users and segments. The basic idea is that an important user is the one who comments and participates the discussion in an important segment, and an important segment is the one with many important users leave their comments. Namely, a segment with many in-links from important users is considered as important clips, and a user with many out-links to important clips is considered as an expert user. The definition of the importance of users and segments can be stated by the following equations; the importance of users contributes to the importance of segments, and importance of segments defines the importance of users.

$$w(u_i) = \sum_{\forall j \in E(u_i)} w(s_j) \tag{2}$$

$$w(s_j) = \sum_{\forall i \in E(u_j)} w(u_i) \tag{3}$$

The details about the computation are shown in Algorithm 2.

Algorithm 2 Mutual Enforcement Scheme for Estimating Users' and Segments' Importance Scores

Input: A set of users U and a set of augmented segments S
Output: A set of weighted users and a set of weighted augmented segments
  for $u \in U$ and $s \in S$ **do**
    initial $w(u)$=1 and $w(s)$=1
  end for
  while $w(u_i)^t - w(u_i)^{t+1} \geq \epsilon$ **do**
    for $i \in U$ **do**
      $w(u_i) = \sum_{\forall j \in E(u_i)} w(s_j)$;
    end for
    for $j \in S$ **do**
      $w(s_j) = \sum_{\forall i \in E(C_j)} w(u_i)$;
    end for
    for $i \in U$ **do**
      $w(u_i) = w(u_i)/Max(U)$;
    end for
    for $j \in S$ **do**
      $w(s_j) = w(s_j)/Max(S)$;
    end for
  end while

Once the computation converges, each clip and user now can be labeled a weight, indicating the importance of a user and a clip. We then respectively extract *top-k* users and clips. The clips are reported as highlights. And, the top-k users and their comments are then employed for annotating the extracted highlights, which were introduced in the next subsection.

3.5 Highlight Semantic Annotation Scheme

At this stage, an extracted highlight and a set of texts related to the highlight are generated by the last stage. At this stage, our goal is to annotate semantics about the extracted highlight. As proposed in [12], a highlight annotation scheme is proposed to work as follows. First, for a given set of comments that belongs to a highlight, the terms related to players, such as name or nickname, and the terms related to play events, such as goals or fouls, are extracted based on a given lexicon containing all player information and event information. Afterward, the results are further organized into two sets: player information set and event information set. As an example, in Figure 4, we show an example for the output of this step, where the player information set contains player names, e.g., Suzuki Ichiro or Jeter, and the event information set contains the events, e.g., Home Runs, mentioned in the comments.

Afterward, the TF-IDF scheme is then employed to weight terms in the following manner. One can treat each augmented segment as a document and the augmented segments of a sport game as a collection of documents. With such a viewpoint, the inverse document frequency weight for the terms shown in a game can be computed. Also, we can compute the frequency of the term in the extracted highlight. As a result, we can label each term in the extracted highlight a weight, and the top-k weighted terms in the two sets are served as the output for annotating an extracted highlight.

However, we find that directly employing all comments shown in an extracted highlight suffers from the effectiveness problem. As mentioned, there are about 230 comments in average for per highlight extracted by either using threshold-based scheme or our proposed scheme. For such concern, our idea is to leverage the result produced by mutual enforcement weighting scheme; namely, we propose to employ comments from expert users only rather than employing all comments, as incorporating all comments are noisy and error-prone. Accordingly, we propose to use the comments from top-k users for highlight semantic annotation, as we believe they will provide information with higher quality.

In summary, we propose to use the comments from expert users as a basis for semantic annotation. More specifically, for a given highlight and the associated comments, we extract only the comments from top-k users as a data source to go through the TF-IDF term weighting and extract terms with high weights to annotate highlights.

Fig. 4 Player Set and Play Set

4 Performance Evaluation

In this section, we report the experiment results for the proposed scheme. In Subsection 4.1, we first introduce the data sets for performance validation, and in Subsection 4.2, we introduce the schemes compared in the experiments and provide an overview of the experiment results. In Subsection 4.4 and Subsection 4.3, we report the comparison results for highlight extraction and highlight annotation, respectively.

4.1 Dataset

We validate the performance of the compared scheme by two sport game types, baseball game and soccer game from 2014 FIFA World Cup and 2016 Challenge Match for Professional Baseball Leagues between Taiwan and Japan. We employ these diverse data sets to validate the robustness of the proposed scheme over different sport games.

Table 1 summarizes the information about each sport game, where the first column shows the game information, the second column shows the teams in the games, the third the match results, the fourth the total number of comments, and the last the data source.

4.2 Compared Schemes

There are two schemes implemented and compared in our study.

- **TBS, Threshold-based Scheme** We implement the scheme introduced in Subsection 3.3, a variant scheme proposed by [12], as a representative of the state-of-the-art. The default values of the parameters for the threshold-based scheme are set to $x = 1.5$, $\alpha = 0.7$, and $w = 60$ seconds, as conducted in [12].
- **MEWS, Mutual Enforcement Weighting Scheme** We also implement the scheme introduced in Subsection 3.4, which extracts highlights by the proposed weighting scheme based on the weighting scheme over the mutual enforcement model.

Table 1 Dataset

Games	Teams	Match Result	Num. of Comments	Data Sources
2014 World Cup				
Semifinals	Netherland v.s. Argentina	2:4	2880 / 8096	[20]
Championship Game	Germany v.s. Argentina	1:0	5043 / 15198	[21]
2016 Taiwan v.s. Japan		5		
Baseball Challenge Match[1]	Taiwan v.s. Japan	3:9	4677 / 12953	[22]

4.3 Evaluation on Highlight Extraction

In this subsection, we validate the performance on highlight extraction for the compared schemes. As a performance overview, we employ MOS (Mean Opinion Score) [23] metric, a subject quality of experience (QoE) measure commonly used for video, audio, and audiovisual quality evaluation, to validate the effectiveness of the compared scheme. The MOS is the arithmetic mean of individual opinion scores, which is computed by the following equation:

$$MOS = \frac{\sum_{n=0}^{N} R_n}{N} \tag{4}$$

, where R_n is the opinion score given by a participant and N is the number of the participants in the evaluation. In our experiments, the opinion score is set to range from 1 (worst) to 5 (best). We invite 30 participants to watch the highlights produced by the compared schemes, and then ask them to give the opinion scores about the effectiveness of the sport games. The performance result is summarized in Table 2. One can see that our scheme provides an average improvement rate of 20% ($= \frac{2.88-2.4}{2.4}$) over the threshold-based scheme, which demonstrates the effectiveness of the proposed scheme. Furthermore, in the experiments, we validate the performance over three games (the baseball game between Taiwan and Japan, the soccer game between Netherland v.s. Argentina, the soccer game between Germay v.s. Argentina). We also show the average MOS scores for the games in Table 2. We see that, for the game between Taiwan and Japan and the game between Netherland and Argentina, our scheme significantly outperforms the threshold-based game. One may notice that for the game between Germay and Argentina soccer game, there is no significant difference between the compared schemes. We find the reason for this result is that the game remain ties with 0:0 until the only goal being scored by Mario Götze in extra time. As there is no significant events in regulation time (most of interesting plays take places in the extra time), no significant signals are observed from social media until the extra time of the game, and therefore the two schemes report similar results, i.e., all highlights are clips in the extra time.

Table 2 MOS Results on Highlight Extraction Comparison

Games	TBS	MEWS
Taiwan v.s. Japan Baseball Challenge Match, NAGOYA	2.019	**2.68**
2014 FIFA Netherland v.s. Argentina	1.733	**2.518**
2014 FIFA Germany v.s. Argentina	3.448	**3.467**
Average MOS	2.4	2.88

4.4 Evaluation on Highlight Annotation

In this section, we report the experiment results for the performance in the highlight annotation. In this set of experiments, we mainly compare the effectiveness on the highlight annotation for the schemes. In the experiments, we compare the following schemes:

- **TBS** In this scheme, the highlights are extracted using the threshold-based scheme and are annotated based on all the comments leaved in the life span of an highlight.
- **MEWS+ALL** In this scheme, the highlights are extracted using the proposed mutual enforcement weighting scheme and are annotated based on the comments from all users with an highlight.
- **MEWS+Expert** In this scheme, the highlights are extracted using the proposed mutual enforcement weighting scheme and are annotated based on only the comments from expert users.

In this set of experiments, we compare the MOS metric and the hit rates of the compared schemes. In Table 3, we report the MOS results of the compared scheme. One can observe that our scheme and its variants outperform the threshold-based scheme and provide an average improvement rate of 16% ($\frac{4.333-3.733}{3.733}$).

Table 3 MOS Results on Highlight Annotation Comparison

Games	TBS	MEWS+Expert	MEWS+ALL
Taiwan v.s. Japan Baseball Challenge Match, NAGOYA	3.4	**4.6**	4.4
2014 FIFA Netherland v.s. Argentina	3.6	**3.8**	3.4
2014 FIFA Germany v.s. Argentina	4.2	**4.6**	4
Average MOS	3.733	**4.333**	3.933

Table 4 Hit Rates on Highlight Annotation Comparison

Games	TBS	MEWS+Expert	MEWS+ALL
Taiwan v.s. Japan Baseball Challenge Match, NAGOYA	0.663	**0.740**	**0.740**
2014 FIFA Netherland v.s. Argentina	**0.733**	**0.733**	0.657
2014 FIFA Germany v.s. Argentina	0.493	**0.860**	0.733
Average Hit Rates	0.630	**0.761**	0.710

In addition, we also report the hit rates of the compared schemes. In this set of experiments, the schemes are set to report Top-3 player name and Top-3 event name. Namely, there are six terms reported by each scheme. The hit rate is defined and computed by the ratio of how many terms match the plots in a detected highlight to being another index for effectiveness measurement. In Table 4 we report the hit rates of the compared schemes. Our schemes again provide an average improvement rate of 20.7% ($\frac{0.761-0.630}{0.630}$).

5 Conclusion

In this paper, we discover the utility of user-generated social texts on social media platform for annotating the semantics of broadcast sport videos. Our goal is to detect important moments and to annotate the moments with semantics. Such goal is essential to sport video summarization and personalized retrieval. In this paper, we present an novel approach for sports video semantic annotation and event detection by mining social media contents. We use soccer game videos and baseball game videos as our test data set. The evaluation over real data shows the effectiveness of the proposed framework.

References

1. T. Sakaki, M. Okazaki, Y. Matsuo, in *Proceedings of the 19th International Conference on World Wide Web* (ACM, New York, NY, USA, 2010), WWW '10, pp. 851–860
2. A. Signorini, A.M. Segre, P.M. Polgreen, PLoS One **6**(5) (2011)
3. E. Aramaki, S. Maskawa, M. Morita. Twitter catches the flu: Detecting influenza epidemics using twitter (2011)
4. B. Pan, Y. Zheng, D. Wilkie, C. Shahabi, in *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (ACM, New York, NY, USA, 2013), SIGSPATIAL'13, pp. 344–353
5. J. Lanagan, A.F. Smeaton, Artificial Intelligence **29**(2), 542 (2011)
6. K. Doman, T. Tomita, I. Ide, D. Deguchi, H. Murase, in *Proceedings of the 22Nd ACM International Conference on Multimedia* (ACM, New York, NY, USA, 2014), MM '14, pp. 949–952
7. Y. Rui, A. Gupta, A. Acero, Proc. eighth ACM Int. pp. 105–115 (2000)
8. Y. Gong, L.T. Sin, C.H. Chuan, H. Zhang, M. Sakauchi, in *Proceedings of the International Conference on Multimedia Computing and Systems* (1995), pp. 167–174
9. D. Zhang, S.F. Chang, in *Proceedings of the Tenth ACM International Conference on Multimedia* (ACM, New York, NY, USA, 2002), MULTIMEDIA '02, pp. 315–318
10. C. Xu, Y.F. Zhang, G. Zhu, Y. Rui, H. Lu, Q. Huang, IEEE Transactions on Multimedia **10**(7), 1342 (2008)
11. N. Nitta, N. Babaguchi, Proc. 8th Int. Work. Multimed. Inf. Syst. '02 pp. 110–116 (2002)
12. L.C. Hsieh, C.W. Lee, T.H. Chiu, W. Hsu, Proc. - IEEE Int. Conf. Multimed. Expo pp. 949–954 (2012)

13. J. Lanagan, A.F. Smeaton, Artif. Intell. **29**(2), 542 (2011)
14. J. Nichols, J. Mahmud, C. Drews, Proc. 2012 ACM International Conference on Intelligent User Interfaces pp. 189–198 (2012)
15. Z. Wang, J. Yu, Y. He, T. Guan, Multimed. Tools Appl. **73**(1), 519 (2014)
16. C. Xu, J. Wang, K. Wan, Y. Li, L. Duan, Proc. 14th Annu. ACM Int. Conf. Multimed. - Multimed. '06 p. 221 (2006)
17. A. Ekin, A.M. Tekalp, R. Mehrotra, IEEE Trans. Image Process. **12**(7), 796 (2003)
18. J. Assfalg, M. Bertini, A.D. Bimbo, W. Nunziati, P. Pala, IEEE Int. Conf. Multimed. Expo (ICME '02) **1**, 825 (2002)
19. F. Sulser, I. Giangreco, H. Schuldt, Proc. 2014 Int. ACM Work. Crowdsourcing Multimed. - CrowdMM '14 pp. 63–68 (2014)
20. 2014 fifa word cup ned 0-0 arg (SF) (2014). URL http://www.ptt.cc/bbs/WorldCup/M.1404933803.A.AE0.html
21. 2014 fifa world cup ger 1-0 arg final (2014). URL http://www.ptt.cc/bbs/WorldCup/M.1405276044.A.502.html
22. (2016). URL https://www.ptt.cc/bbs/Baseball/M.1457254681.A.086.html
23. ITU-T, *ITU-T Recommendation P.85 : A Method for Subjective Performance Assessment of the Quality of Speech Output Devices* (International Telecommunication Union, 1994)