Sentiment Analysis of Wimbledon Tweets

Priyanka Sinha
Tata Consultancy Services Limited
Ecospace 1B, New Town, Rajarhat
Kolkata 700156, India
priyanka27.s@tcs.com

Anirban Dutta Choudhury
Tata Consultancy Services Limited
Ecospace 1B, New Town, Rajarhat
Kolkata 700156, India
anirban.duttachoudhury@tcs.com

Amit Kumar Agrawal
Tata Consultancy Services Limited
Ecospace 1B, New Town, Rajarhat
Kolkata 700156, India
amitk.agrawal@tcs.com

ABSTRACT
Annotating videos in the absence of textual metadata is a major challenge as it involves complex image and video analytics, which is often error prone. However, if the video is a live coverage of an event, time correlated textual feed about the same event can act as a valuable source of aid for such annotation. Popular real time microblog streams like Twitter feeds can be an ideal source of such textual information. In this paper we explore the possibility of such correlation with the sentiment analysis of a set of tweets of the Roger Federer and Novak “Nole” Djokovic semi finals match at Wimbledon 2012.

Categories and Subject Descriptors
I.2.7 [**Natural Language Processing**]: Language Parsing and Understanding; H.1.2 [**User/Machine Systems**]: Human Information Processing

General Terms
Experimentation

Keywords
Twitter, Wimbeldon, Sentiment Analysis, TV

1. INTRODUCTION
Sentiment analysis or opinion mining is an automatic analysis of unstructured text to determine the sentiment expressed in the text such as the polarity of a sentence as either positive or negative.

One way to understand the sentiment of people viewing or experiencing the event is to analyze the video feed from TV or web hosting sites like Youtube. It is a challenging computationally hard problem especially without any helping text. Sentiment annotation in videos at finer granularity is not a much explored area. Sentiment annotations of a live video can be leveraged to enable targeted advertisement. However, there is problem with annotation quality due to the fact that manual video annotation is tedious and time consuming process whereas automated supervised video annotation is very limited in its coverage (i.e. incomplete and sometimes wrong). [7] is an example of how human crowdsourcing is one of the possible ways to annotate videos.

Based on our experiments detailed in ”Our Approach” section, we observe that there exists a correlation between sentiment analysis on tweets and live coverage of video in real time of a popular event, i.e., the Wimbledon semi final match between Roger Federer and Novak Djokovic. This correlation is observed on this particular event and the confirmation of generality of the phenomenon is part of our future work. We have been successful in doing text analysis on microposts for sentiment analysis of a live event with respect to participants in that event in real time which has not been done before. We are proposing a novel approach for automatic sentiment annotation of live coverage of videos related to events affecting mass at large such as politics, natural calamities, sports etc. For a widely well known event with a large number of stakeholders, it is generally seen that the traffic on Twitter is huge with lots of people tweeting about it.

2. RELATED WORK
[8, 6] have demonstrated that twitter based sentiment analysis can be used for closely predicting political election results. However the approach is limited in temporal correlation because the political event (i.e., gold standard) is covered using various news flashes. It is not as fine grained and accurate as capturing the video of the unfolding of political events as they appear on either TV or web.

[5] is similar to our work in that they do discover named entities in tweets and micro-events for live events, but it is a different text mining task than sentiment analysis of the event.

3. OUR APPROACH
Tweets [2] have a maximum length of 140 characters. ”Come on Federer! #Wimbledon” is an example tweet. RT is an acronym for retweet. @ is used to mention a twitter user name. # is used to represent a hashtag. http://bit.ly/9K4n9p is a shortened URL linking to external content.

Published as part of the #Microposts2014 Workshop proceedings, available online as CEUR Vol-1141 (http://ceur-ws.org/Vol-1141)

#Microposts2014, April 7th, 2014, Seoul, Korea.

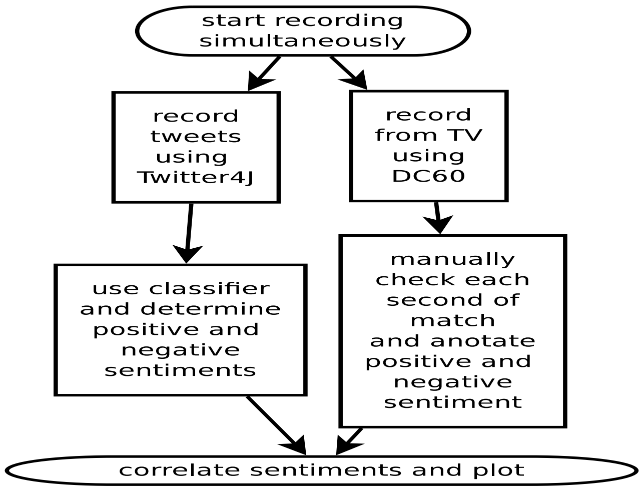

Figure 1: Flowchart for analyzing tweets and correlating results with TV video

Figure 1 depicts the steps we take to find the correlation between manually annotated sentiments of video and tweets. We set up one linux desktop to capture tweets using [4, 3]. Since Twitter uses OAuth 2.0 as authentication to connect to its API, we created an application and generated valid oauth token and secret online for use directly without the OAuth handshake. We grabbed the tweets for "Wimbeldon" keyword during the live telecast. In order to capture the TV video of the live telecast of wimbledon semi final match between "Roger Federer" and "Novak Djokovic", we tuned the Tata Sky set top box to Star Sports and attached a usb [1] to it connecting to a linux laptop. We used mencoder on the linux laptop with settings of aac for audio, h.264 for video and mp4 as the mux. The tweets and the video capture were started almost simultaneously thereby synchronizing the starting timestamps.

Three people manually annotated the video in two column format where the first column was "time in seconds since start of match" and second column was either "1" for positive sentiment for Roger or "2" for positive sentiment for Novak. Majority voting was taken to create the ground truth. We use supervised text classifiers such as Naive Bayes on tweets for sentiment polarity detection. We trained the classifier using part of the tweets which we manually annotated. Finally we used the sentiments derived from analysis on tweets and compared them to the ground truth. We found that the tweet sentiments were correlated with the video, and the time lag between video telecast and tweet was negligible.

4. CONCLUSION AND FUTURE WORK

When game sentiment is towards a particular player, advertisements endorsed by that player can be shown. We can also split the game into parts and get real time summarization of the game sentiment upto that point or within a time span.If intensity of sentiment is used then we can detect peaks of sentiments towards players as well and can tag best moments in the game as well.

Futuristic applications include allowing V-Commerce on live events like popular fashion shows where positive sentiment towards a participant of a video can inform the backend to adjust load towards possible increase in volume of incoming purchases.

Manual annotation to obtain the gold standard from video is a tedious task but gives an accurate understanding of the events sentiment. Video emotional analysis can be used to augment this process.

We understand that since this analysis has been done on one event, similar analysis on more popular real life events would help. Future work would also involve better techniques of sentiment analysis taking into account the short and noisy nature of tweets. Identification and treatment of languages other than english would help for certain events such as the Japanese tweets for Fukushima earthquake. Based on the correlation that we find between tweets and live events in the form of video, we are motivated enough to create a system where automatic annotation of live coverage of an event will take place using sentiment derived from tweets for the same event.

5. ACKNOWLEDGEMENTS

We would like to thank Chirabrata Bhaumik and Avik Ghose in helping us set up the video capture of the Wimbledon semi-finals which gave us our gold standard. We would also like to thank the workshop reviewers for their comments.

6. REFERENCES

[1] Dc 60. http://easycap.blogspot.in/p/easycap-dc60.html.

[2] Twitter. https://dev.twitter.com/docs/platform-objects/tweets.

[3] Twitter streaming api. https://dev.twitter.com/docs/api/1.1/post/statuses/filter.

[4] Twitter4j. http://twitter4j.org/en/index.html.

[5] S. Choudhury and J. G. Breslin. Extracting semantic entities and events from sports tweets. In *Making Sense of Microposts': Big Things Come in Small Packages: co-located with the 8th Extended Semantic Web Conference, ESWC2011*, Heraklion, Crete, May 2011.

[6] T. Rao and S. Srivastava. Analyzing stock market movements using twitter sentiment analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)*, ASONAM '12, pages 119–123, Washington, DC, USA, 2012. IEEE Computer Society.

[7] C. Vondrick, D. Patterson, and D. Ramanan. Efficiently scaling up crowdsourced video annotation. *International Journal of Computer Vision*, 101(1):184–204, 2013.

[8] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time twitter sentiment analysis of 2012 u.s. presidential election cycle. In *Proceedings of the ACL 2012 System Demonstrations*, ACL '12, pages 115–120, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.