Abstract

Recently, twitter users are leveraged to detect social and physical events such as festivals and traffic jam at real time. Real time event detection and summarization from Cricket sports is the process of detecting events such as boundary at real time from live Cricket tweet stream as soon as event happens and generating a quick game summary. This is an interesting, yet a complex problem. Because of the need for rapid detection of sports events and for the generation of a concise summary from huge volume of tweets for Cricket enthusiasts. In this paper, a novel framework is proposed for detecting key events from live Cricket tweets and for generating a game summary using the crawled tweets. Feature vectors of live tweets are created using incremental TF-IDF representation and tweet clusters are discovered using Locality Sensitive Hashing (LSH) where the post rate of each cluster determines the key event. A key event is recognized from that cluster using our domain specific event lexicon. Then, important moments from the crawled tweets are computed by identifying the spikes in the tweets volume. Top-k tweets from each moment are selected by ranking tweets on top-k words.
Representative tweets from top-k tweets are identified using Jaccard similarity. The evaluation on 2017 IPL T20 Cricket live tweets using ROC measure shows that the proposed incremental TF-IDF based LSH approach detects key events with nearly 95% true positive rate and around 5% false positive rate. The proposed game summarization algorithm generates summaries which are readable and competitive to human tailored summaries.

**References**


**Index Terms**

Computer Science Pattern Recognition

**Keywords**
Event Detection, Incremental TF-IDF, Locality Sensitive Hashing, Live Sports Tweets, Event Summarization