

Timelines as Summaries of Popular Scheduled Events

Omar Alonso
Microsoft Corp.
omalonso@microsoft.com

Kyle Shiells
Microsoft Corp.
kshiells@microsoft.com

ABSTRACT

Known events that are scheduled in advance, such as popular sports games, usually get a lot of attention from the public. Communications media like TV, radio, and newspapers will report the salient aspects of such events live or post-hoc for general consumption. However, certain actions, facts, and opinions would likely be omitted from those objective summaries. Our approach is to construct a particular game's timeline in such a way that it can be used as a quick summary of the main events that happened along with popular subjective and opinionated items that the public inject. Peaks in the volume of posts discussing the event reflect both objectively recognizable events in the game -- in the sports example, a change in score -- and subjective events such as a referee making a call fans disagree with. In this work, we introduce a novel timeline design that captures a more complete story of the event by placing the volume of Twitter posts alongside keywords that are driving the additional traffic. We demonstrate our approach using events of major international social impact from the World Cup 2010 and evaluate against professional liveblog coverage of the same events.

Categories and Subject Descriptors

H.4 [Information Systems Applications]: Miscellaneous; H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

Keywords

Timelines, sports, summaries, World Cup, soccer, football, rugby.

1. INTRODUCTION

Microblogging sites like Twitter have gained tremendous attention as platforms to disseminate information about ongoing events. One of the major uses of Twitter is to discuss events in real time. As users join a conversation with their friends and the Twitter community, they generate huge amounts of time-stamped text narrating and discussing the ongoing event. The character limit and time-sensitivity of the content lead people to publish so quickly that only live blogging, a curated form of microblogging, can allow the professional media to keep pace with Twitter in instantaneous coverage.

In the particular case of sports events, users tweet to express their opinions, joys, and frustrations as the game progresses, as well as to report the objectively observable events of the game. The aggregation of the subjective content seems useful as one can see microblogging as another channel to produce and consume

polarized and personal information, in contrast to what one can expect from a professional transmission via TV, radio, or even a live blog. The advantages of scheduled events for the analysis we present are that an audience is established and prepared for discussion from the very beginning of the event and that the start and end times are clearly defined. Media game reports tend to be very objective and highlight certain aspects of the contest but usually miss other facts that a news consumer might also like to read about. Producing a compact summary of a sports game using Twitter data that not only covers the obvious events but also augments them with the crowd's reactions may provide a more complete and satisfying way to represent the full narrative of a game to users.

One can think of our approach as something similar to a *global bar atmosphere*, where fans from both teams playing are carrying on conversations simultaneously alongside other sports enthusiasts, who may not be invested in the specific teams playing but are there to enjoy the sport for its own sake. All the members of the crowd watch the game together and react immediately. Fluctuations in the volume of Twitter traffic can be thought of as differences in the audible volume of conversation in a bar. When a goal is scored, many people react at once and this combines into a roar that can confirm, even to an uninformed observer, that something important must have just happened. The rules of the game define the most salient events (e.g. goal, full-time) and fans will echo those outcomes along with subjective reactions to the events. In addition, fans might respond to a particularly clever play, or (as we will show in an example later) may respond out of proportion to one goal because it represents a milestone for the player who scored it -- an event that would likely be mentioned by a newscaster, but requires contextual knowledge beyond the game itself. How can we capture both the objective and subjective elements of a game and combine them in a visualization of how that game unfolded over time?

Our method is to mine the tweets produced in the online conversation around the event and provide an aggregated summary of the game along a well-defined timeline, with the aim of capturing the most salient parts of the virtual bar activity. Note that we are not interested in sentiment analysis and do not use formal approaches to event or topic detection. Rather, our focus is on a precise timeline that presents the well known facts annotated with the most prominent keywords used by the crowd at a given time, where events can be observed as peaks with associated keywords but are not otherwise separate from the ongoing discussion. We believe this offers an interesting middle ground between the narrative continuity of an editorial summary, as would be presented by a TV broadcast or newspaper, and the huge volume of diverse text that various members of the crowd share in microblogs.

