Twitter, Gambling and Time Sensitive Information

Abstract

One suggested use of 'Big Data' has been prediction markets; whether that be predicting the stock exchange, health outbreaks, or box office revenues. This paper considers another prediction market; WTA Tennis, and considers how the use of Twitter by professional sportswomen can provide information to gamblers and gambling operators, enabling them to take advantage of the information before it becomes widely known. To do so, it is necessary to consider how incoming tweets can be profiled to ascertain firstly whether they contain information which may impact upon a current or future prediction market, and secondly the reliability of information contained within them. Through such profiling, relevant tweets can be made available to decision makers more quickly than current manually operated services which provide information for the gambling industry.

Keywords

Twitter; Big Data; Analysis; Gambling

Twitter & Prediction Markets

Twitter has been used as a tool to attempt to predict a range of activities, including heath outbreaks (Ritterman et al., 2009; Polgreen et al., 2007) , Box office revenues (Asur & Huberman, 2010), the Stock Exchange (Zhang et al., 2010), and elections (Tumasjan et al., 2011) . Whilst predictions of health could be considered for the public good, there are also of course revenue considerations, not least for health insurers and those companies producing potential vaccines. In other cases, there is not even an attempt at altruistic purposes, with the sole intention being to out-perform traditional market indicators, in order to return either direct or indirect revenue for those processing the data, or their employers.

Based on standard Twitter analysis methodologies (Bruns & Burgess, 2011, 2012) I analyzed 217,560 tweets that matched the 'WTA' keyword between April and July 2012, with a particular focus on tweets discussing injuries or form. I compared the content of these tweets, their timing, and their distribution by others (e.g. retweets by journalists) to movement in betting markets for future matches and tournaments. In many cases, the timing of the tweet preceded a significant change in the prediction market. I also considered the role of 'amplifiers'; television or internet sports personalities or writers who, as well as breaking their own news, re-tweet that of colleagues.

Value of Information in the Gambling Industry

There are clear advantages to the use of twitter by gambling industry participants to gather information on player injuries, team selection and other factors. While in the US context much of this is performed under the guise of fantasy sports information (which is league sanctioned), in other markets the significance of information for gambling is acknowledged, both by those distributing the information and through platforms which aggregate it, such as *TennisForm.com.*

The value of 'inside information' is clear. While NBA referee Tim Donaghy received prominent attention for potentially altering the outcome of matches he was in charge of, his account also acknowledges (Donaghy, p. 4) he had an "inside advantage because of [his] access to pregame meetings. It was common for my fellow referees to voice their opinions about who they expected to win on a given night. Those opinions were often based on their knowledge of confidential inside information pertaining to players and teams, such as injury reports unknown to the general public".

Many professional sporting leagues, including all of the major United States sports and Australia's NRL and AFL have strict reporting requirements for injuries. Whilst the NFL, in line with their anti-gambling stance, asserts in their bylaws that the purpose of injury reports are to "tell the opposing

coach of the injury status of his players so that each coach can plan strategies for the game" (Harding, 2006), the use of such information in gambling and fantasy sports is undeniable. Borghesi et al. (2009), in a study considering Arena Football, noted how player injury status in the sport meant that "diligent bettors may, at times, have an information advantage over bookmakers". The significance of this information for predicting sporting events is then clear.

Market Adjustments

Whilst mainstream betting sports such as the NFL, NBA, College Football, College Basketball and Soccer would have their markets impacted by injury developments, this current study focuses on tennis, a sport where the entire outcome of a match, and a significant impact on tournament markets, is dependent on a single player. A large number of players on both the ATP and WTA tour have twitter accounts, and these accounts are frequently used to discuss niggling injuries, upset stomachs, flu symptoms, or to simply announce a withdrawal from an upcoming tournament. Tweets such as ""Mica has recovered. We have trained together. I believe everything is OK." (Nenad Zimonjic about Michael Llodra, 19 March 2012), or "Update: feeling back to normal and practicing" (Vania King, 17 March 2012) will inevitably have an impact on betting markets while not being stories which would attract mainstream press coverage.

Additionally, the range of markets currently offered, particularly online, means that any tweets of this kind would have an impact on prediction markets, whether it be the market for a match the same day, the outright market for an upcoming tournament, or a long term performance market, such as the odds for a player to win a tournament in the current calendar year.

Whilst sites such as *Tennis Form* aggregate data on player injuries (at a cost of 50 Euros per month), the ability to filter and access this information in real time is key to obtaining an advantage on the remainder of the market. In this way, the same question is raised as with the stock exchange, with health, or with crisis management; how do you filter the information and place the result in front of somebody who can decide what to do with it.

Selecting Twitter Information

Whilst analysis of past events is useful, it is more relevant to consider how such tweets could be isolated for future use. While the volume of tweets directly from WTA approved player accounts is relatively low, at between 50 and 200 tweets per day, the broader conversation (which includes coaches, agents, training partners etc) on the WTA keyword amounted to a further 2000/day, and this would rise significantly if other keywords or accounts were added to the data set. Within these tweets is useful information, but also substantial rumor and misinformation. Indeed, the network graph (Figure 1) shows a large number of players, but also news organizations and fans prominent in the conversation, whose guesses at injury status may not be reliable.

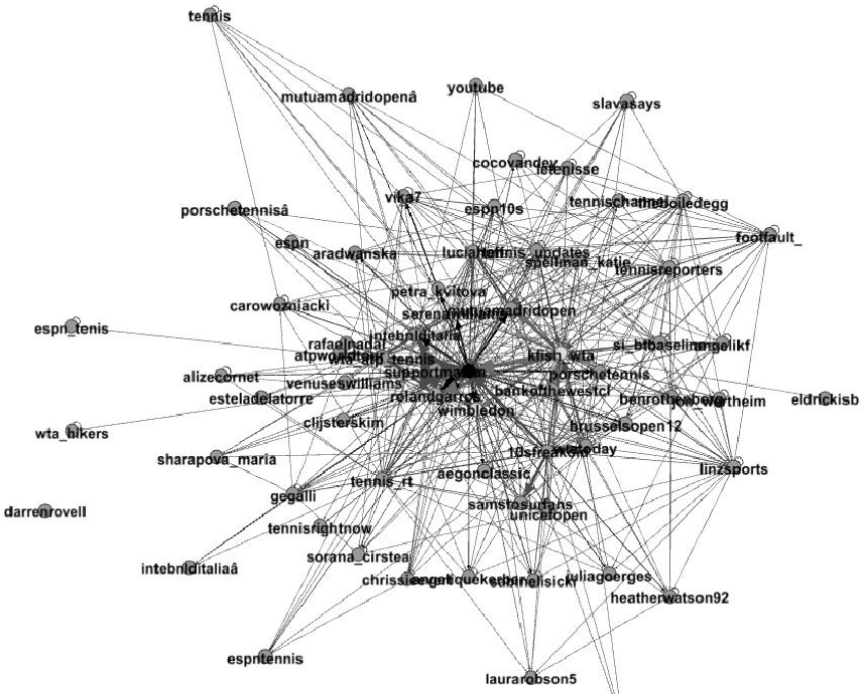

Figure 1: Twitter Network -- #WTA hashtag, April-July 2012

However, through an analysis of past tweets, there appear to be key indicators as to the reliability of information, which, whilst requiring manual judgment to set up, may enable automated filtering. Those tweets directly from the players account can be generally assumed reliable (unless the player has a history of misinformation). A mapping of the players immediate social and professional network (who have twitter accounts) can be produced through an analysis of their @ replies on Twitter, and thus a reliability score can also be assigned to these accounts, in proportion to the volume of tweets between them and the player. Finally, external sources such as travelling reporters and photographers, as well as event staff for each WTA tournament must be considered; a process which involves both research and qualitative judgment. In all cases, a retweet by one of the 'significant' accounts has value; if the retweet is of another trusted source this would serve to amplify the information, whilst a retweet of a third party would raise attention, but require further validation.

Conclusion

Bookmakers and gamblers alike are charged with evaluating each piece of information, the source, and the potential impact on the betting market; a service with a value, and a significant potential for profit. The stakeholders, their requirements and a detailed understanding of the domain is necessary to consider how developments in automatically ranking tweets could be applied to any given set of information, whether a betting market, health scare or crisis situation.

References

Asur, S., & Huberman, B. (2010). Predicting the Future with Social Media. In *Proceedings of the ACM Conference on Web Intelligence, 2010.* Retrieved from http://arxiv.org/pdf/1003.5699.pdf.

Borghesi, R., Paul, R., & Weinbach, A.P. (2009). Market Frictions and Overpriced Favorites: Evidence from Arena Football. In *Applied Economics Letters,* 16(9): 903-906

Bruns, A. & Burgess, J. (2011). Tools. Retrieved from http://mappingonlinepublics.net/resources/.

Bruns, A. & Burgess, J. (2012). Researching News Discussion on Twitter: New Methodologies. Journalism Studies, 13.5-6.

Donaghy, T. (2010). Personal Foul: A First-Person Account of the Scandal that Rocked the NBA. Sarasota, FL: Four Daughters LLC.

Harding, Casey N. (2006). *Nickel and Dimed: North Carolina court blocks Carolina Panthers' attempt to avoid payment of workers' compensation benefits to injured athletes.* In 28 N.C. Cent. L.J. 241

Polgreen, P.M., Nelson, F.D., Neumann, G.R., & Weinstein, R.A. (2007). Use of Prediction Markets to Forecast Infectious Disease Activity. In *Clinical Infectious Diseases,* 44(2):272-279.

Ritterman, J., Osborne, M. & Klein, E. (2009). Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic. In *Workshop on Mining Social Media*. Retrieved from http://www.christopia.net/data/school/2011/Fall/social-media-mining/project_proposal/sources/ritterman-2009.pdf

Tumasjan, A., Sprenger, T.O., Sandner, P.G., & Welpe, I.M. (2011). Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. In *Social Science Computer Review*, 29(4):402-418.

Zhang, X., Fuehres, H., & Gloor, P.A. (2010). Predicting Stock Market Indicators Through Twitter- I Hope it is not as bad as I fear. In *Collaborative Innovations Networks Conference (COINs)*.