Big Data Analysis of Major Elite Sport Events in The Netherlands. Case Studies of the 2017 UEFA Women's Championship and the Rabo EuroHockey Championships 2017

Hover, Paul[1]; Mol, Peter-Jan[2] and Wassenaar, Kim[2]
1: Mulier Institute, Netherlands, The; 2: Knowledge Centre for Sport Netherlands
p.hover@mulierinstituut.nl

Aim
The growing usage of the internet, mobile devices and social media by an increasing number of people offers opportunities for investigating the web to explore and track interests, attitudes and preferences, including sport (events) related content. A substantial part of Dutch citizens are involved in following sport events, as spectators and via (online) media, which can also be witnessed in sport event related content in social media, blogs and other platforms, producing big data (Newcom.nl, 2018). Despite considerable governmental investments in the organisation of major sport events and measuring its social impact and a rapid development of social media management software, research on big data related to major sport events is very scarce. It was for this reason this explorative study on big data related to the 2017 UEFA Women's Championship and The Rabo EuroHockey Championships 2017, both held in the Netherlands, was developed.

The aim of this study is threefold. Firstly, to assess the development of the level of online interest of these two elite sport events, measured in number of posts and their potential reach. Secondly, to evaluate the sentiment of these events by assessing the development of the ratio of positive, neutral and negative online content. In the third place, to reflect on the method and results and to appoint strengths and weaknesses.

Theoretical Background and Literature Review
Technology and the growing levels of internet penetration have significantly transformed the traditional ways of marketing communication, also regarding sport events. In social media consumers are simultaneously the initiators and recipients of sport event information exchanges. Measuring the level of interest and perception of major sport events using online generated content to is worldwide in its infancy (Yu & Wang, 2015; Burch et al., 2017).

Research Design and Data Analysis
Two methods were applied, namely quantitative text analysis (QTA) and sentiment analysis (SA). With QTA large amounts of text are automatically processed. With SA online content is automatically categorised in positive, neutral and negative content. Ten types of online sources were considered: Twitter, Facebook, Instagram, news websites, blogs, forums, Google+, YouTube, LinkedIn and Pinterest.

It was an ex post analysis where per event three time slots were considered. These were the period of four weeks in the run up to the event, the event period and the four weeks after the event. The social media management software of Coosto, one of the leading software packages in the Dutch market, was applied for tracing, processing and analysing data.
The search queries were created via a snowball method. Each search started with the name of the event. Based on the trending topics page of the software package, which displays terms related to the search, the search was stepped up until a saturation level was reached and the addition of a new search term no longer generated a significant amount of new content.

Results and Discussion

In the period studied - the event itself and the four weeks before and after - 172,000 messages were sent about the 2017 UEFA Women's Championship, with a potential of 945 million contact moments. 4 percent of the messages was negative, 69 percent neutral and 27 percent positive. The Rabo EuroHockey Championships 2017 showed similar findings on a smaller scale, despite this being a double gender event: 31,000 messages were posted, with a potential of 230 million contact moments; 4% of the messages was negative, 71% neutral and 25% positive. The excellent performances of the Dutch teams – they all won – undoubtedly skewed the results in a positive way.

Strengths of this type of analysis include the fact that there are no costs involved for data collection (e.g. Xin et al., 2010). Moreover, online posts are often spontaneous and real-time reactions and emotions (e.g. Yu & Wang, 2015) and the data offer good opportunities for monitoring over time. The people who post messages on the internet are not representative for the population, which is a noteworthy weakness. In addition, the software recognizes no sarcasm, no irony and no incorrect information.

Conclusion and Implications

With the high internet penetration and numerous platforms to communicate, major sport events result in a continuously stream of data during the event, but also before and after it, which can be construed as a form of collective wisdom (Asur & Huberman, 2010). Despite being still rather general, without providing much in-depth information, the data offer interesting insights which are a valuable addition to more common types of research, like an economic impact analysis. Our results provide sufficient grounds to be positive about the capability of QTA and SA to become an even better method for capturing the level of online interest of sport events and to evaluate the sentiment of these events.

References

Asur, S. & Huberman, B.A. (2010). "Predicting the Future with Social Media", 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, 2010, pp. 492-499. doi: 10.1109/WI-IAT.2010.63.

Burch, L.M., Billings, A.C. & Zimmerman, M.H. (2017). Comparing American soccer dialogues: sociale media commentary Surrounding the 2014 US men's and 2015 US women's World Cup teams, *Sport in Society*, DOI: 10.1080/17430437.2017.1284811.

Newcom.nl (2018). Nationale Social Media Onderzoek 2018. Het grootste trendonderzoek van Nederland naar het gebruik en verwachtingen van social media #NSMO. Amsterdam/Enschede: Newcom Research & Consultancy B.V.

Xin J., Gallagher A. & Cao, L. (2010). The wisdom of social multimedia. In: Proceedings of ACM multimedia 2010 international conference, Firenze, 25–29 October, pp. 1.235–1.244.

Yu, Y. & Wang, X. (2015). World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets, *Computers in Human Behavior*, 48 (2015), pp. 392–400. Elsevier.