Analysis of Olympic Tweets from Tokyo 2020

Naman Bhargava, Jordan Rodrigues
Georgia Institute of Technology
North Ave NW
Atlanta, GA 30332
nbhargava9@gatech.edu, jrodrigues30@gatech.edu

Abstract

This work seeks to understand how and why individuals utilized Twitter to engage with the 2020 Tokyo Olympics via tweet data with the hashtag *#Tokyo2020* from [start date] to [end date]. We extend on previous work that analyzed the 2010, 2012, and 2016 Olympics, as well as studies on Twitter engagement for the NBA and the 2014 World Cup. In particular, we aim to answer [rq1], [rq2], and [rq3] through the usage of [tech1], [tech2], and [tech3]

Introduction

The Olympic Games are one of the world's most universal athletic competitions, drawing more than 200 competing nations [#1]. Digital engagement with the Olympics has also been increasing drastically, with more than 100 million unique users participating for Tokyo 2020, doubling the amount of users from the previous Summer Olympics [#2]. Of the online platforms, Twitter captured a majority share of Olympics discussion, generating more than 75 billion impressions during Rio 2016 (compared to Facebook's 1.5 billion interactions) [#3]. As online engagement has increased, so too has the diversity of discussion that takes places during the event; individuals converse about the introduction of new sports, the performance of athletes, and celebrate the success of their nations. However, not all Twitter discussion is positive. Following the emergence of COVID-19, anti-asian racism and xenophobia rose drastically [#4], prompting significant controversial discussion around the postponement of Tokyo 2020, as well as COVID protocols for the games. Negative sentiment is also often displayed against competing nations and individual athletes as well.

We seek to conduct a holistic overview of Twitter discussion during the Tokyo 2020 Olympics. This will include studying [rq1], [rq2], and [rq3]. These questions will be answered via the *Tokyo 2020 Olympics Tweets Dataset* [#5], which contains a sample of tweets using the hashtag *#Tokyo2020* between 07/24/2021 and 07/27/2021, the beginning of the Games.

What is the motivating problem for your research project? What are the research questions to be examined for the project? If you are contributing a dataset, what do you envision the output of your work to look like? Who can use this dataset, and on what kind of problems would this dataset be ideal? If you are contributing analysis, how are these findings theoretically significant or broadly important for society? -0.5 pts, introduction does not clearly describe a motivating problem -1 pt, introduction does not mention the research goals/questions clearly

Related Work

Other works have looked at Olympics related discussions through tweets on Twitter. Sentiment analysis was conducted on a Rio de Janeiro Olympic tweets dataset released by (Vertalka, Kassens-Noor, and Wilson 2019). This dataset included over 21 million tweets with location data, language, and tweet content. This sentiment analysis was done in (Kassens-Noor, Vertalka, and Wilson 2019), finding tweets about host cities were positive, while tweets about the IOC were negative. The study scored tweet sentiment by taking the difference between positive and negative key words matched from the Mohammad and Turney's (2010) lexicon. This method meant only English language tweets could be scored. This work was important to show the positive effects on host city image perception from hosting the Olympics on a global population. Our work focuses more on understanding different topics and sentiments around those topics of discussion, rather than just focussing on host cities.

General sentiment analysis and engagement is explored by (Gruzd, Doiron, and Mai 2011) during the Winter 2010 Olympics. This paper found positive tweets were forwarded more and negative users were more passionate. The study calculated sentiment using the SentiStrength system. The work poses future questions on how people engage or cluster around topics or tweet tones, and how social network impacts engagement. Our work focuses on understanding the topics that have the highest engagement.

Methods

Methods (3 pts) 1-3 paragraphs explaining your research methodology. What data will you use? What model? What statistical tests will you run? -1.5 pt, the data pipeline is not clearly or specifically described -1.5 pt, the computational model / statistical analysis is not clearly or specifically described

Timeline

1 short paragraph describing a weekly timeline for your project and task responsibilities by team members. By when do you hope to have acquired your data? When will you build or select the statistical / computational models and when will you run them? When will you finalize the main analyses and plots and paper writing? -0.5 pts, basic project milestones are not identified and assigned to team members -0.5 pts, rough dates for completing basic project milestones have not been identified

Page Breaks

For your final camera ready copy, you must not use any page break commands. References must flow directly after the text without breaks. Note that some conferences require references to be on a separate page during the review process. AAAI Press, however, does not require this condition for the final paper.

References. The references section should be labeled "References" and should appear at the very end of the paper (don't end the paper with references, and then put a figure by itself on the last page). A sample list of references is given later on in these instructions. Please use a consistent format for references. Poorly prepared or sloppy references reflect badly on the quality of your paper and your research. Please prepare complete and accurate citations.

Formatted bibliographies should look like the following examples.

Book with Multiple Authors
Engelmore, R., and Morgan, A. eds. 1986. *Blackboard Systems.* Reading, Mass.: Addison-Wesley.

Journal Article
Robinson, A. L. 1980a. New Ways to Make Microcircuits Smaller. *Science* 208: 1019–1026.

Magazine Article
Hasling, D. W.; Clancey, W. J.; and Rennels, G. R. 1983. Strategic Explanations in Consultation. *The International Journal of Man-Machine Studies* 20(1): 3–19.

Proceedings Paper Published by a Society
Clancey, W. J. 1983. Communication, Simulation, and Intelligent Agents: Implications of Personal Intelligent Machines for Medical Education. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence,* 556–560. Menlo Park, Calif.: International Joint Conferences on Artificial Intelligence, Inc.

Proceedings Paper Published by a Press or Publisher
Clancey, W. J. 1984. Classification Problem Solving. In *Proceedings of the Fourth National Conference on Artificial Intelligence,* 49–54. Menlo Park, Calif.: AAAI Press.

University Technical Report
Rice, J. 1986. Poligon: A System for Parallel Problem Solving, Technical Report, KSL-86-19, Dept. of Computer Science, Stanford Univ.

Dissertation or Thesis
Clancey, W. J. 1979. Transfer of Rule-Based Expertise through a Tutorial Dialogue. Ph.D. diss., Dept. of Computer Science, Stanford Univ., Stanford, Calif.

Forthcoming Publication
Clancey, W. J. 2021. The Engineering of Qualitative Models. Forthcoming.

References

Gruzd, A.; Doiron, S.; and Mai, P. 2011. Is Happiness Contagious Online? A Case of Twitter and the 2010 Winter Olympics. In *2011 44th Hawaii International Conference on System Sciences*, 1–9. doi:10.1109/HICSS.2011.259.

Kassens-Noor, E.; Vertalka, J.; and Wilson, M. 2019. Good games, bad host? Using big data to measure public attention and imagery of the Olympic Games. *Cities* 90: 229–236. ISSN 0264-2751. doi:https://doi.org/10.1016/j.cities.2019.02.009. URL https://www.sciencedirect.com/science/article/pii/S026427511830307X.

Knuth, D. E. 1984. Literate Programming. *The Computer Journal* 27(2): 97–111.

Knuth, D. E. 1986. *The TEX Book*. Addison-Wesley Professional.

Lamport, L. 1994. *LATEX: a Document Preparation System*. Massachusetts: Addison Wesley, 2 edition.

Vertalka, J. J.; Kassens-Noor, E.; and Wilson, M. 2019. Data on sentiments and emotions of olympic-themed tweets. *Data in Brief* 24: 103869. ISSN 2352-3409. doi:https://doi.org/10.1016/j.dib.2019.103869. URL https://www.sciencedirect.com/science/article/pii/S2352340919302203.
1986; 1984; 1994