

Topic Entropy : How to discriminate general topics from biased topics on social media

Takeshi Sakaki
Hottolink, The Univ. of Tokyo
6, Yonbancho, Chiyoda-ku
Tokyo, Japan
t.sakaki@hottolink.co.jp

Fujio Toriumi
The University of Tokyo
7-3-1, Hongo, Bunkyo-ku
Tokyo, Japan
tori@sys.t.u-tokyo.ac.jp

Mizuki Sakae
Hottolink.
6, Yonbancho, Chiyoda-ku
Tokyo, Japan
s.mizuki@hottolink.co.jp

ABSTRACT

Along with the popularization of social media, many people try to observe trend topics by mining social media data in various fields. People often use the volume of social media posts as a quantitative indicator for each topic. However, the volume of posts does not seem to be an appropriate indicator because of the "the generality of topics": some trend topics are diffused broadly and thinly, others are diffused narrowly and densely. Emphasizing biased topics sometimes causes possibly severe problems: exaggerating the voice of vocal minority, diffusing inaccurate news with a catchy title.

As described herein, we propose a method to measure how biased topics on social media quantitatively, "Topic Entropy", by analyzing user clusters that are involved in each topic. In evaluation experiments, we analyze three burst cases including natural disasters and flaming. Thereby, we uncover some big topics that are raised by various user groups, other big topics are brought up by a few user groups.

Keywords

burst phenomena, topic clustering, social media mining, user community detection

APPENDIX

A. DATASET DESCRIPTION

A.1 McDonald's alien substance mixture case

Search Queries

"McDonald" "Mac" (in Japanese)

Abstract

The case that alien substances were discovered in products of McDonald's in succession from December to 2014 through January, 2015.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

World Wide Web Conference '17 Perth, Australia

© 2017 ACM. ISBN 978-1-4503-2138-9.

DOI: 10.1145/1235

Time period

18th Dec. 2014 - 14th Jan. 2014

A.2 ALS Ice Bucket Challenge in Japan

Search Queries

"ALS", "Ice bucket", "iced water"(in Japanese)

Abstract

Social Movements to choose whether dump a bucket of iced water on his/her head or donate to the American ALS association. This activity started from the U.S. 2014. This became a topic because many celebrities involved this activity, while some people asserted critical opinions about it

Time period

2014/8/1-2014/9/2

A.3 Plagiarism scandal of Olympic Emblem

Search Queries

"emblem", "Kenjiro Sano" (in Japanese)

Abstract

The case that it was suspected that the emblem of the 2020 Tokyo Olympics by Mr. Kenjiro Sano might be plagiarism and the emblem was finally withdrawn.

Time period

2015/7/26-2016/9/21