2014 年 6 月 2 日

●バージニア工科大研究者、ツイート・データ使用する際の 倫理ガイドラインを提案

【Virginia Tech, 2014/06/02】

　バージニア工科大バージニア・バイオインフォマティクス・インスティテュートの研究者は、ツイッターからのデータを倫理的に収集・利用するためのガイドラインを提案した。

　ツイッターは他の SNS と違い、リアルタイムのデータ・ストリーミングが可能な複数の API を提供しており、データ・ストリームからのメタデータにはユーザーの現在地を正確に示す情報が含まれることもあり、情報的価値が高い反面、慎重な取り扱いが求められる。

　今回、バージニア工科大の研究チームが発表したガイドラインは、科学分野の研究にツイッターのデータを使う際の指針を提案し、議論を深めようとするもの。米国消費者プライバシー権利章典をモデルとする同ガイドラインは

1) 調査の目的、手法、データ取扱い方法を公開し、簡単に閲覧できるようにする
2) ツイートの匿名性を守る
3) プライベート、あるいは削除されたツイートは調査対象から除外する
4) ツイートデータを他のソースからの個人情報と統合しない
5) プライバシーや匿名性を損なう可能性がある調査方式については施設内倫理委員会と協議する
6) ツイートの文脈上の関連性を尊重する、
等を提示している。

　現在、ツイッターのデータ調査を行う場合はライブストリームされているツイートの約 1%、もしくは過去に遡ってユーザー1 人あたり最大 3200 件のツイートにアクセスできる。またツイッターは今年 2 月、同社の持つ過去の公共データに全てアクセスできるパイロット・プログラムも開始している。

（参考）本件報道記事
Researchers propose tactics for ethical use of Twitter data

Tweets are only 140 characters, but amass enough of them, and researchers can deduce a great deal about subjects ranging from disease outbreaks to social unrest. Researchers with the Virginia Bioinformatics Institute of Virginia Tech

have proposed guidelines to make sure data mined from Twitter data is obtained and used ethically.

Twitter has emerged as one of the most popular social networks in the world. Billions of tweets present an unfiltered view of public opinions and experiences. Researchers can deduce a lot from those 140 characters — from tracking disease outbreaks to predicting social unrest.

Unlike other social networking sites, Twitter provides multiple application programming interfaces that allow for real-time data streaming. Metadata from the stream can even include geolocations that pinpoint the exact geographic coordinates of a user.

The vast amount of data available from tweets is extremely beneficial for researchers, but handling that information requires an ethical framework. Caitlin Rivers, a doctoral student in the Genetics, Bioinformatics, and Computational Biology Program, along with Bryan Lewis, a research scientist in the Network Dynamics and Simulation Science Laboratory at the Virginia Bioinformatics Institute of Virginia Tech, have proposed a set of ethical guidelines for scientific researchers using Twitter data. Their opinion article first appeared in February in F1000 Research, an open science journal for life scientists.

Currently researchers can access a random sample of approximately 1 percent of all live-streamed tweets or retrospectively gather up to 3,200 tweets from a single user. In 2010, Twitter donated its entire historical record of tweets to the U.S. Library of Congress.

In February 2014, Twitter introduced a pilot program that would give researchers full access to Twitter's public and historical data. The variety of ways researchers can utilize Twitter as a research tool is likely to continue to grow.

By viewing online spaces with the same regard as physical spaces, there are social norms regarding privacy. However, Twitter users often do not expect their movements will be tracked or their personal information recorded. Even in public places, common standards of privacy are observed.

"Why should online spaces like Twitter not be subject to those same standards?" said Rivers, a Department of Defense Science, Mathematics, and Research for Transformation scholar.

Researchers interested in user-centric studies look at the Twitter activity of individuals, but doing so raises potential privacy concerns. Twitter data is public information, but many individuals claim a reasonable expectation of

privacy. Even if a user's assumptions on privacy are incorrect, there are presumptions that because something is shared publicly, the user consents to it being used for research.

"Twitter is an amazing data source for scientists, but it's also a potential privacy risk," said Rivers. "Our guidelines are simple ways for researchers to respect the privacy of Twitter users while still conducting valuable research."

Modeled after the U.S. Consumer Privacy Bill of Rights, Rivers and Lewis outline six guidelines for the ethical use of Twitter data:

-- Make objectives, methodologies, and data handling practices transparent and easily accessible.

-- Protect the anonymity of tweet authors by not publishing identifiable information without consent.

-- Honor Twitter users' efforts to control their personal data by omitting private and deleted tweets.

-- Do not use tweet data to aggregate personal information from other sources.

-- Work collaboratively with the Institutional Review Board for study designs that may compromise privacy and anonymity.

-- Respect the context in which a tweet was sent.

Lewis points out these tactics are "not to preclude using data in this manner, but only to encourage researchers to adhere to existing ethical standards." These proposed guidelines are meant to inspire an open discussion among the research community. "We are at a point in digital research where the technical advances are outpacing standards of ethical practice," Rivers said.

Source:
http://www.alphagalileo.org/ViewItem.aspx?ItemId=142402&CultureCode=en

以　上