Assignment 3

Take Home Exam

Student Name: Wenhao CHEN
Student Number: 13372949

Introduction

In this paper, I will introduce the challenges of three main challenges when the analyst predicts the data through social media data. And it will provide some ideas to solve the challenges, and at the end of the paper, I will list the ethical problems and comparison of ethical codes.

One of the reasons for social media communication is the opportunity to receive or create and share public messages in a reduced, ubiquitous way. The huge increase in the use of social media has led to an increase in data accumulation, which is called social media big data.

Challenge 1: How to collect the data

According to the model of The Social Media Analytics Framework (Stieglitz et al., 2014; Stieglitz & Dang-Xuan, 2013), Data is required to perform different operations. Fan and Gordon (2014) proposed a three-step process, which is, capture, understand and present. Based on the three-step process, Stieglitz et al. (2014) also propose a framework for social media analytics. How to properly capture social media data about the topic 'vote' has become one of the challenges.

The solution

After determining the steps, we analysed the characteristics of social media data. There are many similarities between social media data and big data. Therefore, I drew on many kinds of literature analysing big data. One solution is to collect data on different platforms by building different models and algorithms. For example, character limitation requires a corresponding algorithm and model, which is consistent with the data extraction of Twitter social media. For the Facebook platform, the staff needs to build another model and algorithm to discover and track new data and collect them.

In the general case of data analysis, most algorithms and models are universal, but in the data analysis of social media, we should establish different data analysis models and algorithms, corresponding to the framework for social media analytics proposed by Stieglitz et al (2014). In the enhanced discovery stage, before tracking data.

Challenges 2: How to save the data

As we know, social media data is a huge amount and it must find a good way to save them. McAfee, A., Brynjolfsson, E. (2012) put forward three "V" problems and challenges of big data, one of which is how to store big data. Storage of social media data is also A big challenge. The relational database management system is a common data analysis, but according to different social media data collection, traditional structured query language can

not meet the needs of big data.

The solution

Databases that use unstructured query statement languages are more efficient for social media data than traditional database storage systems. Does not apply SQL as a query language, the establishment does not contain fixed format data type classification, suitable for various types of data. NoSQL is a method used to build a database, which replaces the traditional relational database system.

Challenges 3: How to find the correct data

As Google's director of research, Peter Norvig, puts it: "We don't have better algorithms. We just have more data." In social media data, there is data that includes almost any topic from different devices, such as mobile phones, social networks, GPS, etc., and data types may also include more data types like Numbers, words, slang, and different languages. Even in life, each of us produces a statistic called step count. The available data is often unstructured and complex, and it is a difficult challenge to find and apply more efficient data after the NoSQL database has been set up and stored in it.

The solution

Discarded missing value

The Missing Value of the data cannot be unified, because the data types are diverse, and most of the data will hardly contain all types. The existence of Missing Value is necessary, and how to deal with the Missing Value is one of the challenges. Usually, in data analysis, data with missing values are thrown away directly after the data type is selected.

Filling missing value

Fill in missing data values. Filling in missing data values is a better method than deleting data because it is a good method to fill in missing values through statistical analysis of the median or mode of data

Evaluation algorithm

The existence of missing values always affects social media data and data analysis. How to better fill in missing values is challenging. Not all data can be filled in with mode or missing values. As a member of a voting institution, in many cases, the data they are faced with is not only yes and no. The staff should establish corresponding data prediction methods through different data types, and fill in the missing value data of unconventional classes utilizing data prediction.

Ethical Issue and Social Consequence

In most international codes of ethics, data collection should protect the privacy of individuals, as mentioned by IEEE, ACS, and ACM. Although we legally collect user information (the user agrees to the terms and conditions), the collection and use of social media data is still a violation of the code of ethics. In the process of forecasting data, users may feel that

their privacy is violated because we retrieve the corresponding data without the user's knowledge, and this process may not prompt users about sensitive topics. Privacy infringement of social media data may upset the public and hinder the use of social media. Social media platforms (such as Facebook and Twitter) will also be affected.

Reference

1. Andrew McAfee & Erik Brynjolfsson, Oct 2012, *Big Data: The Management Revolution,* Viewed 8th Oct 2019 <https://hbr.org/2012/10/big-data-the-management-revolution>

2. ACS, 2019, *Australian Computer Society*, Viewed 8th Oct 2019, <https://www.acs.org.au/>

3. ACM, 2019, *Association for Computing Machinery*, Viewed 8th Oct 2019, <https://www.acm.org/>

4. D. boyd & K. Crawford, 2012, *Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon Information Communication and Society,* 15 (5) (2012), pp. 662-679,

5. E. Hargittai, 2015, *Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites The ANNALS of the American Academy of Political and Social Science,* pp. 63-76

6. GMI Blogger, August 2017, *4 Challenges Of Using Social Media Data For Analytics,* Digital Analytics & Business Intelligence, Viewed 8th Oct 2019, <https://www.globalmediainsight.com/blog/4-challenges-using-social-media-data-analytics/>

7. IEEE, 2019, *IEEE Compliance,* Viewed 8th Oct 2019, <https://www.ieee.org/about/compliance.html>

8. I. Guellil & K. Boukhalfa,2015, *Social big data mining: A survey focused on opinion mining and sentiments analysis 12th international symposium on programming and systems,* pp. 132-141

9. McAfee, A. & Brynjolfsson, E., Oct 2012, *Big Data*. Harvard Business Review, (October) Pages 60-68

10. Sapir Segal, 2019, *The What, Why, and How of Social Media Data*, Viewed 8th Oct 2019, <https://www.oktopost.com/blog/social-media-data/>

11. Stefan Stieglitza & Milad Mirbabaie & Björn Ross & Christoph Neuberger, *Social media analytics – Challenges in topic discovery, data collection, and data preparation,*

Volume 39, April 2018, Pages 156-168