

Understanding society through social computing: Implementations of culture, media, and governance

PROPOSAL

**WebSci'21 | 13th ACM Web Science Conference
Globalisation, Inclusion and the Web in the Context of COVID**

21-25 June 2021

Hosted by the University of Southampton, UK, delivered online
<https://websci21.webscience.org/>

WORKSHOP

**Understanding society through social computing: Implementations of culture,
media, and governance**

Topic:

This workshop focuses on implementation of machine learning methods such as text mining and image analysis techniques to study social issues. Over the past few decades, the explosive increase of unstructured data, e.g., text and image, offers a great opportunity for social scientists to understand complex social and cultural phenomena. Assisted with machine learning techniques, contents such as images and texts that were difficult to capture in the past, can be digitized and analyzed, which promotes the emerging field of social computing. For example, by using word vector technology and topic clustering method in machine learning, scholars can capture the deep cultural, social, and even gender implications inside a large amount of movie and literature texts. Thus, through social computing, shadowy concept such as “culture”, “society”, or “information cocoon” become a tangible phenomenon. Furthermore, analyzing online texts, images and visual information can help scholars to understand how social media mobilizes collective action, and explore how text features of online opinions influence government’s response, which could be vital for understanding society in digital age. In addition, social computing can also improve governance, such as predicting the rate of domestic violence and introducing policies to intervene, and evaluating the dynamic process between measures adopted by local government and negative social sentiment in the context of the COVID pandemic.

1 Unspeakable domestic violence: the real domestic violence rate predicted by machine learning

Professor Yunsong Chen, Nanjing University

Professor Senhu Wang, National University of Singapore

Abstract: The traditional social investigation method is often difficult to obtain the real rate of domestic violence since the victims of domestic violence are often reluctant to disclose the fact of their own domestic violence. It results in the underestimation of the proportion of investigation. This study uses machine learning algorithms to predict the real incidence of domestic violence among domestic partners in China based on the data of The Third China Women's social status survey. Through the comparison of five unbalanced sample processing methods and six machine learning algorithms, we ensure the optimal validity of the prediction. The predicted results show that the real rates of "physical violence," "verbal violence," and "cold violence" among partners are about 7.31%, 14.34%, and 22.22% respectively, while the original statistics are 4.05%, 11.21%, and 17.95%, which underestimate the three kinds of domestic violence. In particular, on cold violence, there are differences between the models before and after the prediction in gender and in urban and rural. It indicates that machine learning effectively adjusts the bias of the original incomplete samples and greatly improves the overall goodness of fit of the model.

Keywords: domestic violence, machine learning, social prediction, missing value

2 Literary Destination Familiarity and Inbound Tourism: Evidence from China

Associate Professor Fei Yan, Tsinghua University

Associate Professor Guangye He, Nanjing University

Abstract: Destination familiarity is an important non-economic determinant of tourists' destination choice that has not been adequately studied. This study posits a literary dimension to the concept of destination familiarity—that is, the extent to which tourists have gained familiarity with a given destination through literature—and seeks to investigate the impact of this form of familiarity on inbound tourism to China. Employing the English fiction dataset of the Google Books corpus, the New York Times annotated corpus, and the Time magazine corpus, we construct two types of destination familiarity based on literary texts: affection-based destination familiarity and knowledge-based destination familiarity. The results from dynamic panel estimation (1994–2004) demonstrate that the higher the degree of affection-based destination familiarity with a province in the previous year, the larger the number of inbound tourists the following year. Examining the influence of literature and its consumption on tourism activities sheds light on the dynamics of sustainable tourism

development in emerging markets.

Keywords: Inbound tourism, tourist, destination familiarity, China, Google Books

3 The Geometry of Information Cocoon: Analyzing the Cultural Space with Word Embedding Models

Associate Professor ChengJun Wang, Nanjing University

Abstract: Accompanied by the rapid development of digital media, the threat of information cocoon has become a significant issue in our society. The purpose of this study is threefold: first, to provide a geometric framework of information cocoon; second, to examine the existence of information cocoon in the daily use of digital media; and third, to investigate the relation between information cocoon and social class. We construct the cultural space with word embedding models among three large-scale datasets of digital media use. Our analysis reveals that information cocoons widely exist in the daily use of digital media across mobile apps, mobile reading and computer use. Moreover, people of lower social class have a higher probability of being stuck in the information cocoon filled with the content of entertainment. In contrast, the people of higher social class have more capability to stride over the constraints of information cocoon.

Keywords: Information cocoon, Cultural space, Word embedding, Cultural consumption, Social class

4 Graphic Knowledge Production of Chinese Advertising Ephemera in Early 19th Century

Associate Professor Jing Chen, Nanjing University

Abstract: Researches have shown that commercial ads are historically embedded text/images that began appearing in Chinese lithographic print news media in the late 19th century. Commercial centers, or “treaty ports,” participated in a new, worldwide, ad industry that developed slogans, sophisticated cartoon drawings, innovative fonts and syntax to sell machine-made, branded, commercially exchanged large and small commodities. Ads became a powerful vehicle for circulating desirable commodities in a modernist commercial culture. For us the significant point is that black and white newspaper ads are “[m]inor transient documents of everyday life,” ephemera, which we see as a graphic epistemology. The ads or “graphs” forward axiological ideas like “modern things are clean” or “people are mammals,” for instance, but they do it wordlessly. (Drucker) Social theorists and historians in the first third of the 20th century agreed commercial ephemera were modern and had value. Barlow has argued

that ads are part of the modern disciplinary order consisting of psychology, sociology, political science, etc. Supporting generic ads there eventually emerged an entire commercial industry devoted to buying and selling advertising space. Since people determine social value our potential scholarly users can exploit our ad archive to discover not only how 1920s and 1930s media space was sociologically quantified, bought and sold, but also how brilliant impresarios like Morishita Hiroshi, C.P. Ling and Carl Crow creatively pioneered complex ad campaigns that were pedagogical, linguistically innovate and philosophically unprecedented. An immediate question is how modern Chinese language and visual media are modern. We focus on neologisms or calques (new words) to illustrate the integration of social theory and advertising culture. For instance, one of the most important of all the modern words in 20th century Chinese is “society.” “Society” is not a descriptor but a category of experience. It has a long career and its novelty has been established using traditional history methods. But the word is also a part of everyday life in 20th century commercial culture. This research is based the metadating (in Manovich’s words) , text mining and visual analysis of images archived by the Chinese Commercial Advertisement Archive to draw a picture of “modern society” visually and textually to rise more questions like how these mixtures of ads have shaped the everyday recognition of modern life of ordinary people.

5 Awakening or loathing of the self? Women’s Perspectives in Chinese Online BL Fictions

Ph.D. Candidate Wen Ma, Nanjing University

Abstract: BL (Boy’s love) fictions, also known as danmei fictions in Chinese, are stories describing romances between men. Since the 21st century, it has gradually flourished on the Chinese Internet and has increasingly become a popular cultural form. The difference between BL fictions and ordinary gay novels is that most of the authors and readers of BL are women, and it describes women’s fantasy of gays, rather than the real gay community. In this case, we explored these women’s perspectives and the social and cultural background. Based on Jinjiang Literature City (<http://www.jjwxc.net>), the largest online BL fiction platform in China, we adopted the LDA topic model, word vector technology, and Baidu’s sentiment analysis to study 79,668 original BL fictions from 2003 to 2019. We found that the idea of misogyny and feminist consciousness coexisted in these fictions, through which women expressed their yearning for eros and power, but the existing patriarchy still bound their consciousness. Our study creatively applied machine learning technology to conduct content analysis, comprehensively examined the features of numerous texts at a macro level, and finally drew a conclusion conducive to understanding the contradictions of women’s perspectives presented in BL fictions.

Keywords: Boys’ love, Online novel, Feminism, Content analysis, Machine learning

6 A Chinese Tale of Three Regions: A Century's China in Thousands of Films

Ph.D. Candidate Zhuo Chen, Nanjing University

Ph.D. Candidate Guodong Ju, The London School of Economics and Political Science (LSE)

Abstract: In the past century, the film industries of the three regions—mainland China, Taiwan, and Hong Kong SAR, exhibited different characteristics from one cultural root due to their different historical and social backgrounds. Using the Internet movie database (IMDb), we applied big data analysis and machine learning methods to compare the contents, topics, and sentiments of “Image of China” spread by different regions’ films. We also studied the contained historical and cultural background. The findings indicated that, during ups and downs, the three regions seek their subjectivity and strive to connect with the globalized world in films. The macroscopic analysis of large-scale content enabled us to explore the hidden cultural phenomena and reality behind the media and made up for the lack of objectivity of traditional research methods.

Keywords: Film, International communication, Image of China, Machine learning, Big data

7 Does the crying baby always get the milk? An analysis of government responses for online requests

Associate Professor Tianji Cai, University of Macau

Abstract: Selective response of government to its citizens has been one of essential topics in social sciences. Arguably, one of characteristics that differentiate democracy from other types of form of government is its continuing responsiveness to citizens' preferences. Previous studies have shown that not only democratic governments respond to public demand in order to win election, but authoritarian regimes, more or less, also answer to public demands. However, government's response can be very selective.

The Chinese government has been investing resources to build information infrastructure in the past few decades. Online engagement between representatives of local government and citizens has become one of popular channels for individuals to raise concerns, ask questions and express opinions on local affairs. However, except for a few, little research attention has been paid for how online participation influences both government and people in China, especially at local level. Furthermore, as many of studies explored the effect of institutional or social economic factors at macro level

on government response, unfortunately, one thing that has been frequently overlooked is text itself. One may wonder if a scrawling post delays response.

To fill this gap, the current study aims to address whether text features, such as logic structure, semantic connection, and topic relation, influence government's response. To be specific, we want to evaluate if text features of a post could influence the chance and the speed of local government's response, all else being equal. A total of 113,146 posts that cover from July 24th, 2014 to Nov 10th, 2020 were successfully retrieved from the website "Luzhou Wenzheng", which is the official online platform for Luzhou City located in the southeast of Sichuan Province, China. Our results indicate that text features, especially the number of topic and the logic structure within a post significantly influence the speed of response.

8 Agent-Based Modeling of Empire Dynamics

Professor Peng Lv, Central South University

Abstract: The periodic law of the Empire is steady and persist. There seems to be a general director that governs the dynamics process of human society. This paper aims to reveal this mechanism behind the macro-level systemic law in China. In order to make up for the weak aspects of static theoretic research and data mining, a new paradigm of multi-agent system simulation is adopted. From the perspective of social structure and social actors, a multi-agent system (MAS) including tiger (empires' enemy), wolf (ruling class), sheep (peasant class) and grass (agricultural land) is constructed. Therefore, the goal of high-definition back-calculating the 2132-year historical dynamics, from the Qin to Qing Dynasties, can be achieved. Through parameter randomization and genealogy, continuous simulations can be carried out, in order to find the optimal parameter solution with the highest fitness of history. This optimal solution not only matches the span distribution of empires in history, but also precisely corresponds to the order and spans of empires. Although now China has no risk of the periodic law, this study still has far-reaching strategic significance.

Keywords: periodic law of dynasty, rise and fall, multi-agent simulation

9 How Protesters Use Pictures During Mobilization and Why It Matters

Assistant Professor Han Zhang, The Hong Kong University of Science and Technology

Abstract: Protesters often use textual and visual information to mobilize collective action. Yet, past research has mostly relied on text as data to infer protester's framing strategies. Recent scholarship argues that images are more easily processed and might trigger emotional responses more easily than text, but these arguments are not verified by comparing text and image-based mobilization directly. Using the largest

protest event dataset in China, CASM-China (with over 130,000 protest events and over 200,000 social media posts and images associated with protest events) and automated text and image analysis techniques from computer science, we found that there are substantial differences in how protesters use text and images to achieve their goals. Moreover, images in general and key attributes of images contribute to online mobilization more than text. Our research contributes our understanding of how online mobilization is performed in social media age.

10 Growing networks with heterogenous dense communities via friendships

Research Associate Keke Shang, Nanjing University

Abstract: Almost all existing approaches adopt the evolution principle - a key ingredient of the Barabasi-Albert model - to propose network models which evolve multiple communities. However, previous studies neglect variation in community link density which is a natural consequence of the broad range of different social structures in a single social network and which can provide densely connected pathways along which information spreads rapidly. Hence, we take advantage of social transitivity to provide a novel growth model which naturally develops communities spanning a spectrum of link densities. Finally, we confirm that simulated spreading processes using our model match those of online social networks with heterogenous density.

11 PM2.5 exposure and anxiety in China: evidence from the prefectures

Associate Professor Wei Guo, Nanjing University

Abstract:

Background: Anxiety disorders are among the most common mental health concerns today. While numerous factors are known to affect anxiety disorders, the ways in which environmental factors aggravate or mitigate anxiety are not fully understood.

Methods: Baidu is the most widely used search engine in China, and a large amount of data on internet behavior indicates that anxiety is a growing concern. We reviewed the annual Baidu Indices of anxiety-related keywords for cities in China from 2013 to 2018 and constructed anxiety indices. We then employed a two-way fixed effect (FE) model to analyze the relationship between PM2.5 exposure and anxiety at the prefectural level.

Results: The results indicated that there was a significant positive association between PM2.5 and anxiety index. The anxiety index increased by 0.1565258 for every unit increase in the PM2.5 level ($P < 0.05$), which suggested that current PM2.5 levels in China pose a considerable risk to mental health.

Conclusion: The enormous impact of PM2.5 exposure indicates that the macroscopic environment can shape individual mentality and social behavior, and that it can be

extremely destructive in terms of societal mindset.

Keywords: Anxiety, PM2.5, Baidu index, Two-way FE model, China

12 The Evolution of Negative Sentiments Under the Pandemic and Large Scale Lockdown: A Big Data Analysis Based on Twitter and GDELT

Associate Professor Weigang Gong, Wuhan University

Professor Yunsong Chen, Nanjing University

Abstract:

Drawing from multiple sources of online posts, this study aims to explore the pattern of how negative social sentiment changes in the context of the COVID pandemic. In particular, the focus of study is to evaluate the connection between governmental measures (e.g., lockdown) and negative social sentiment, and predict the change of negative social sentiment through machine learning algorithms.

The analysis revealed that almost entire world experienced a similar trend of social sentiment changes during the pandemic, characterized by declining positive sentiment and growing negative sentiment. The trend of growing negative sentiment can be classified into two phrases: the first phase is featured by fear and panic, while depressed, hopeless, and despaired are the main characteristics of the second stage. Results obtained from fixed effects models further showed that measures adopted by government against the pandemic played a critical role in shaping the dynamic of social sentiment. Although enforced social distancing, community confinement and mass quarantine reduce the chance of contagion, negative implications of such measures affecting public mental health are evident, including depression and stress disorder. Besides the findings above, our machine learning model showed high predictive power on negative sentiments, which can be used to evaluate social impact of governmental measures in the future.

作者群：

- 1 陈云松 王森浒 中国家庭暴力的机器学习 新加坡国立 南大社会学 教授
- 2 严飞 贺光烨 文化熟悉度与旅游大数据面板模型 清华社会学 副教授
- 3 王成军 信息茧房 三大社交媒体数据的词向量分析 机器学习 南大新闻传播 副教授
- 4 陈静 民国广告的知识图谱和文化新词 数字人文 南大艺术学院 副教授
- 5 马文 耽美小说 同性恋文学。词向量情感分析等机器学习 南大社会学 博士生
- 6 陈茁 句国栋 一国三影 电影大数据结合词向量 LSE 伦敦政经 社会学 博士生
- 7 蔡天骥 澳门大学 中国政府文本的机器学习 澳门大学 副教授
- 8 吕鹏 历史的帝国模型 仿真 中南大学 教授
- 9 张涵 社会抗议的图像文本机器学习 香港科大 助理教授
- 10 尚可可 社会网络演化的仿真建模 南大 助理研究员
- 11 郭未 pm2.5 和焦虑 面板环境大数据 南大副教授
- 12 龚为纲 陈云松 疫情中的 G20 国家隔离和推特情绪 武大副教授 南大社会学 教授

