# CS 6474/CS 4803 Social Computing: Prediction & Forecasting II

## Munmun De Choudhury

munmund@gatech.edu

Week 14 | November 18, 2019

### Predicting the Future With Social Media

Sitaram Asur Social Computing Lab HP Labs Palo Alto, California Email: sitaram.asur@hp.com Bernardo A. Huberman Social Computing Lab HP Labs Palo Alto, California Email: bernardo.huberman@hp.com

Abstract—In recent years, social media has become ubiquitous and important for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, we demonstrate how social media content can be used to predict real-world outcomes. In particular, we use the chatter from Twitter.com to forecast box-office revenues for movies. We show that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. We further demonstrate how sentiments extracted from Twitter can be further utilized to improve the forecasting power of social media. This paper reports on such a study. Specifically we consider the task of predicting box-office revenues for movies using the chatter from Twitter, one of the fastest growing social networks in the Internet. Twitter <sup>1</sup>, a micro-blogging network, has experienced a burst of popularity in recent months leading to a huge user-base, consisting of several tens of millions of users who actively participate in the creation and propagation of content.

We have focused on movies in this study for two main reasons.

# A Long List of Predictions....

- Predicting the H1N1 flu outbreak using Twitter data
- Predicting the outcome of the 2016 Presidential elections in the US using social media data
- Predicting people's home location from their geolocated social media posts
- Predicting traffic conditions every day during rush hour based on geo-located social media posts
- Predicting people's diurnal moods based on social media activity
- Predicting an individual's mental health state from the language of social media posts

### Treading with caution

Attention to noise, bias, and "provenance" — broadly, where did data arise, what inferences were drawn from the data, and how relevant are those inferences to the present situation?



Photo credit: Peg Skorpinski

#### Artificial Intelligence — The Revolution Hasn't Happened Yet



Artificial Intelligence (AI) is the mantra of the current era. The phrase is intoned by technologists, academicians, journalists and venture capitalists alike. As with many phrases that cross over from technical academic fields into general circulation, there is significant misunderstanding



- So many things can be predicted with social media, but when and what should be predicted and with what goal in mind?
- What are the implications of these predictions in the world and on people?
- When not to predict.

Private Traits and Attributes are Predictable from Digital Records of Human Behavior

# Summary

- Facebook "likes" used to predict a range of highly sensitive personal attributes like ethnicity, religious and political views, intelligence, happiness, parental separation, age and gender.
- 58K users of Facebook who consented to authorize the mypersonality app
- Participants took many sociometric and psychometric tests
- Predictive accuracies were very high for sexual orientation, parental separation, political views, and the openness attribute of Big Five personality scale
  - The algorithms proved 88% accurate for determining male sexuality, 95% accurate in distinguishing African-American from Caucasian-American and 85% for differentiating Republican from Democrat.
  - Christians and Muslims were correctly classified in 82% of cases and relationship status and substance abuse was predicted with an accuracy between 65% and 73%.

#### Your Friends Personalities

Most Like Me

Your Personality Soulmate

Similarity Score: 85.77% (How was this calculated?)

Sofie Jansson

Trait	0	50	100	% (diff.)
0				50% (-)
С				56% (-25%)
E				88% (+13%)
A				56% (-7%)
N				69% (+13%)

Maybe Opposites Attract? Damon Alexander Young Similarity Score: 75.87% (How was this calculated?)

Least Like Me



Trait	0	50	100	% (diff.)
0				94% (+44%)
С				56% (-25%)
E				88% (+13%)
A				69% (+6%)
N				44% (-12%)

Friend's Name	Personality					Similarity	View	View Full	Friend	
riteriu s Name	0	С	Е	А	Ν	Score	Graph	Profile	Status	
You	50%	81%	75%	63%	56%					
Sofie Jansson	50%	56%	88%	56%	69%	86%	82	-	6	
Sara Lee	88%	63%	63%	69%	56%	80%	82	-	$\varkappa$	
Damon Alexander Young	94%	56%	88%	69%	44%	76%	80	-		

# The Pluses First...

- Largest FB study outside of FB
- Provocative findings
  - It is indeed amazing so many (sensitive) things can be predicted
  - Validity of Facebook data as reflective of people's underlying traits and behaviors
  - Possibility of educating people how simple metrics like "likes" can be profusely revealing of deeply personal information
- Data collected with consent (?)
- But at what cost?

## **Class Exercise I**

What are the limitations and concerns of releasing a Facebook app and using it to collect people's data? What type of biases does it introduce?

# Revisiting the Pluses...

- Largest FB study outside of FB
- Provocative findings
  - It is indeed amazing so many (sensitive) things can be predicted
  - Validity of Facebook data as reflective of people's underlying traits and behaviors
  - Possibility of educating people how simple metrics like "likes" can be profusely revealing of deeply personal information
- Data collected with consent (?)
- But at what cost?

### Who benefits from this research?



11 March 2013 Last updated at 15:00 ET

#### Facebook 'likes' predict personality

Sexuality, political leanings and even intelligence can be gleaned from the things you choose to "like" on Facebook, a study suggests.

Researchers at Cambridge University used algorithms to predict religion, politics, race and sexual orientation.

The research, **published in the journal PNAS**, forms surprisingly accurate personal portraits, researchers said.

The findings should "ring alarm bells" for users, privacy campaigners said.

The study used 58,000 volunteers who alongside their Facebook "likes" and demographic information also provided psychometric testing results - designed to highlight personality traits.

The Facebook likes were fed into algorithms and matched with the



What do your Facebook likes say about you?

#### Related Stories

< Share 🛛 🚹 💟 🖹

Facebook changes privacy settings Privacy groups lobby Facebook

#### **Top stories**



Soldier killed in Canada shootings

WHO crisis team holds Ebola talks Sweden could use force in sub hunt U-boat found with ship it sank Total plane crash driver 'got lost'

#### ADVERTISEMENT



#### Patenta



"This research should ring alarm bells for anyone who thinks that privacy settings are the solution to protecting information online. We need to fundamentally re-think how much data we are voluntarily sharing," said Nick Pickles, director of privacy campaign group Big Brother Watch.

"Yet again, it is clear the lack of transparency about how users' data is being used will lead to entirely justified fears about our data being exploited for commercial gain." Search Q

#### **Disrupt Berlin 2019**

Startups

Apps

Gadgets

Videos

Audio

Extra Crunch

Newsletters

Events

Advertise

Crunchbase More

Apple Enterprise Transportation Facebook privacy

### Facebook bans first app since Cambridge Analytica, myPersonality, and suspends hundreds more

Devin Coldewey @techcrunch / 9:14 pm EDT • August 22, 2018



\_

Comment

## **Class Exercise II**

Ability to infer accurately individual traits can have implications in better personalization and search, what are its risks in privacy?

What are the other implications of such inferences?



# Gaydar and the Fallacy of Decontextualized Measurement

Andrew Gelman,<sup>a</sup> Greggor Mattson,<sup>b</sup> Daniel Simpson<sup>c</sup>

a) Columbia University; b) Oberlin College; c) University of Toronto

**Abstract:** Recent media coverage of studies about "gaydar," the supposed ability to detect another's sexual orientation through visual cues, reveal problems in which the ideals of scientific precision strip the context from intrinsically social phenomena. This fallacy of objective measurement, as w <sup>++</sup> term it, leads to nonsensical claims based on the predictive accuracy of statistical significance. We interrogate these gaydar studies' assumption that there is some sort of pure biological measur <sup>++</sup> of perception of sexual orientation. Instead, we argue that the concept of gaydar inherently exist \_\_ within a social context and that this should be recognized when studying it. We use this case as an example of a more general concern about illusory precision in the measurement of social phenomena



PEER-REVIEWED JOURNAL ON THE INTERNET

HOME ABOUT LOGIN REGISTER SEARCH CURRENT ARCHIVES ANNOUNCEMENTS SUBMISSIONS

Home > Volume 19, Number 7 - 7 July 2014 > Tufekci



PEER-REVIEWED JOURNAL ON THE INTERNET

Engineering the public: Big data, surveillance and computational politics

by Zeynep Tufekci

#### Abstract

Digital technologies have given rise to a new combination of big data and computational practices which allow for massive, latent data collection and sophisticated computational modeling, increasing the capacity of those with resources and access to use these tools to carry out highly effective, opaque and unaccountable campaigns of persuasion and social engineering in political, civic and commercial spheres. I examine six intertwined dynamics that pertain to the rise of computational politics: the rise of big data, the shift away from demographics to individualized targeting, the opacity and power of computational modeling, the use of persuasive behavioral science, digital media enabling dynamic real-time experimentation, and the growth of new power brokers who own the data or social media environments. I then examine the consequences of these new mechanisms on the public sphere and political campaigns.

#### Contents

Introduction Engineering the public: From broadcast to the Internet New dynamics of persuasion, surveillance, campaigning and social engineering Consequences and power of big data analytics Discussion and conclusion

### CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar\*, Jeremy Irvin\*, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, Matthew P. Lungren, Andrew Y. Ng

#### We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists.

Chest X-rays are currently the best available method for diagnosing pneumonia, playing a crucial role in clinical care and epidemiological studies. Pneumonia is responsible for more than 1 million hospitalizations and 50,000 deaths per year in the US alone.



# Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.

#### Yilun Wang, Michal Kosinski

Created on: September 07, 2017 | Last edited: October 16, 2017

م 🗆	▲ Page: 1 of 47 - + Automatic Zoom ÷	23 🖨 🚿	>	Download preprint	Downloads: 24829
	DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM F.	ACES		Abstract	<b>9 F</b> in <b>4</b>
1 2	THIS IS A PREPRINT OF THE PEER REVIEWED ARTICLE TO APPEAR IN JO PERSONALITY AND SOCIAL PSYCHOLOGY.	URNAL OF		We show that faces contain much more information about s perceived and interpreted by the human brain. We used dee features from 35,326 facial images. These features were ent	exual orientation than can be ep neural networks to extract ered into a logistic regression
3	THE MOST RECENT VERSION IS AVAILABLE AT https://osf.io/zp79	k/		See more	image, a classifier
5	AUTHOR NOTES ARE AVAILABLE AT: https://goo.gl/9b2aR2	~		Preprint DOI	
6				10.17605/OSF.IO/HV28A	
7 8 9	Deep neural networks are more accurate than humans at detecting sexual orientation images	from facial		License	
10	Yilun Wang, Michal Kosinski			No license 🕨	
11	Graduate School of Business, Stanford University, Stanford, CA94305, U	USA		Disciplines	
12	michalk@stanford.edu			Social and Behavioral Sciences Psychology	
13				_	
15	The study has been approved by the IRB at Stanford University			Tags	
16 17	Citation: Wang, Y., & Kosinski, M. (in press). Deep neural networks are more accurate	ate than		Artificial Intelligence Big Data Computational Social Science Facial recognition Prenatal Hormone Theory Privacy Sexua	Computer Vision I orientation
18	humans at detecting sexual orientation from facial images. Journal of Personali	ty and		Citations	
ang_kosinski.p	df	Download preprint Version: 1	0	APA	worke are more accurate than

Wang, Y., & Kosinski, M. (2017, October 16). Deep neural networks are more accurate than humans at detecting sexual orientation from facial images.. Retrieved from psyarxiv.com/hv28a

# **Class Debate**

Is technology (here predictive models that use social media data) political or neutral?

FOR and AGAINST argument with the given abstracts

Prediction and explanation in social systems

# Summary

- Prediction is becoming central in the study of (online) social systems
- Papers presents three issues that need resolution to be able to derive value out of predictive approaches
- Standards of prediction:
  - Use multiple evaluation metrics
  - Models are evaluated by third parties (e.g., the Netflix prize)
  - Begin with exploratory research move to confirmatory; register research design data etc.
- Limits of prediction
  - Theoretical limit to predictive accuracy
  - Consideration of confounding factors
  - Calibrate expectations; reduce false optimism
- Prediction versus interpretation
  - Prediction and interpretation do not have to be a trade-off
  - Hybrid approach where simple and complex methods are combined and the solution is question driven

# What are the limits of prediction? Can they fail?

### FINAL FINAL

#### BIG DATA

## The Parable of Google Flu: Traps in Big Data Analysis

David Lazer, 1,2\* Ryan Kennedy, 1,3,4 Gary King, 3 Alessandro Vespignani 3,5,6

n February 2013, Google Flu Trends (GFT) made headlines but not for a reason that Google executives or the creators of the flu tracking system would have hoped. Nature reported that GFT was predicting more than double the proportion of doctor visits for influenza-like illness (ILI) than the Centers for Disease Control and Prevention (CDC), which bases its estimates on surveillance reports from laboratories across the United States (1, 2). This happened despite the fact that GFT was built to predict CDC reports. Given that GFT is often held up as an exemplary use of big data (3, 4), what lessons can we draw from this error?

The problems we identify are not limited to GFT. Research on whether search or social media can



Large errors in flu prediction were largely avoidable, which offers lessons for the use of big data.

the algorithm in 2009, and this model has run ever since, with a few changes announced in October 2013 (10, 15).

Although not widely reported until 2013, the new GFT has been persistently overestimating flu prevalence for a much longer time. GFT also missed by a very large margin in the 2011-2012 flu season and has missed high for 100 out of 108 weeks starting with August 2011 (see the graph). These errors are not randomly distributed. For example, last week's errors predict this week's errors (temporal autocorrelation), and the direction and magnitude of error varies with the time of year (seasonality). These patterns mean that GFT overlooks considerable information that could be extracted by traditional

#### Meaningless comparisons lead to false optimism in medical machine learning

Orianna DeMasi<sup>1</sup>, Konrad Kording<sup>2, 3</sup>, and Benjamin Recht<sup>1</sup>

 <sup>1</sup>Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA
<sup>2</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
<sup>3</sup>Department of Neuroscience, University of Pennsylvania, Philadelphia, PA, USA
*odemasi@berkeley.edu, kording@upenn.edu, brecht@berkeley.edu*

July 21, 2017

#### Abstract

A new trend in medicine is the use of algorithms to analyze big datasets, e.g. using everything your phone measures about you for diagnostics or monitoring. However, these algorithms are commonly compared against weak baselines, which may contribute to excessive optimism. To assess how well an algorithm works, scientists typically ask how well its output correlates with medically assigned scores. Here we perform a meta-analysis to quantify how the literature evaluates their algorithms for monitoring mental wellbeing. We find that the bulk of the literature ( $\sim 77\%$ ) uses meaningless comparisons that ignore patient baseline state. For example, having an algorithm that uses phone data to diagnose mood disorders would be useful. However, it is possible to over 80% of the variance of some mood measures in the population by simply guessing that each patient has their own average mood - the patient-specific baseline. Thus, an algorithm that just predicts that our mood is like it usually is can explain the majority of variance, but is, obviously, entirely useless. Comparing to the wrong (population) baseline has a massive effect on the perceived quality of algorithms and produces baseless optimism in the field. To solve this problem we propose "user lift" that reduces these systematic errors in the evaluation of personalized medical monitoring.

n terms that match the propenbut are structurally unrelated, t predict the future, were quite velopers, in fact, report weedal search terms unrelated to the ly correlated to the CDC data, regarding high school baskets should have been a warning lata were overfitting the small ses, a standard concern in data s ad hoc method of throwing search terms failed when GFT nissed the nonseasonal 2009 -H1N1 pandemic (2, 14). In ial version of GFT was part flu winter detector. GFT engineers

Considering the large number of approaches that provide inference on influenza activity (16-19), does this mean that the current version of GFT is not useful? No, greater value can be obtained by combining GFT with other near-real time health data (2, 20). For example, by combining GFT and lagged CDC data, as well as dynamically recalibrating GFT, we can substantially improve on the performance of GFT or the CDC alone (see the chart). This is no substitute for ongoing evaluation and improvement, but, by incorporating this information, GFT could have largely healed itself and would have likely remained out of the headlines.

#### **SCIENCE** VOL 343 14 MARCH 2014

Source: D. Lazer, R. Kennedy, G. King, A. Vespignani. "The Parable of Google Flu: Traps in Big Data Analysis," Science, vol. 343, March 14, 2014

12

## Class Exercise IV

Assess whether in each of the following cases interpretation or prediction (or both) is/are preferred.

- i) How people react on a new product release (e.g., an iphone), as observed on social media
- ii) Whether greater anonymity leads to greater hate speech on social media
- iii) Whether more ads on YouTube videos leads to lesser YouTube use
- iv) Whether exposure to similar ideological content leads to reinforcement of existing ideologies
- v) Assessments of inconvenience of the I85 highway crash, based on people's social media activity