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

Munmun De Choudhury

munmund@gatech.edu

Examples of many successes…

Science

Current Issue   First release papers   Archive   About ⌄   Submit manuscript

f   y   in   reddit   wechat   mail

Prediction and explanation in social systems

JAKE M. HOFMAN , AMIT SHARMA , AND , DUNCAN J. WATTS   Authors Info & Affiliations

🔔   🔖   ❞   🔒 GET ACCESS

Abstract

Historically, social scientists have sought out explanations of human and social phenomena that provide interpretable causal mechanisms, while often ignoring their predictive accuracy. We argue that the increasingly computational nature of social science is beginning to reverse this traditional bias against prediction; however, it has also highlighted three important issues that require resolution. First, current practices for evaluating predictions must be better standardized. Second, theoretical limits to predictive accuracy in complex social systems must be better

Incomplete history of cascade prediction

Who	Predicting	Features	Metric	Conclusion
HongD 10	Is item retweeted?	Topic Models	F1=0.47	Better than baseline
JendersKN 13	Will item reach some size T?	Content	F1>0.9	High accuracy
TanLP 14	Which of two does better?	Wording	Accu=65.6%	Computers are OK
ChengADKL 14	Will cascade double?	Temporal	AUC=0.88	Predictable
Lerman, Yang, Petrovic, Romero, Kupavskii, Ma, Weng, Zhao, Yu, etc				

Progress?

All of this work examines a different question with a different measure of success, evaluated on a different subset of data, making it difficult to assess overall progress[1]

---

[1]http://hunch.net/?p=22

Predicting success on Twitter?

Bakshy, Hofman, Mason, Watts (2011):

How viral will my tweet be?

"Cascades are unpredictable!"

Reasons behind the inconsistencies

Meaningless comparisons lead to false optimism in medical machine learning

Orianna DeMasi[1], Konrad Kording[2, 3], and Benjamin Recht[1]

[1]Department of Electrical Engineering and Computer Sciences, University of California Berkeley, Berkeley, CA, USA
[2]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA
[3]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.

Exploring limits to prediction in complex social systems
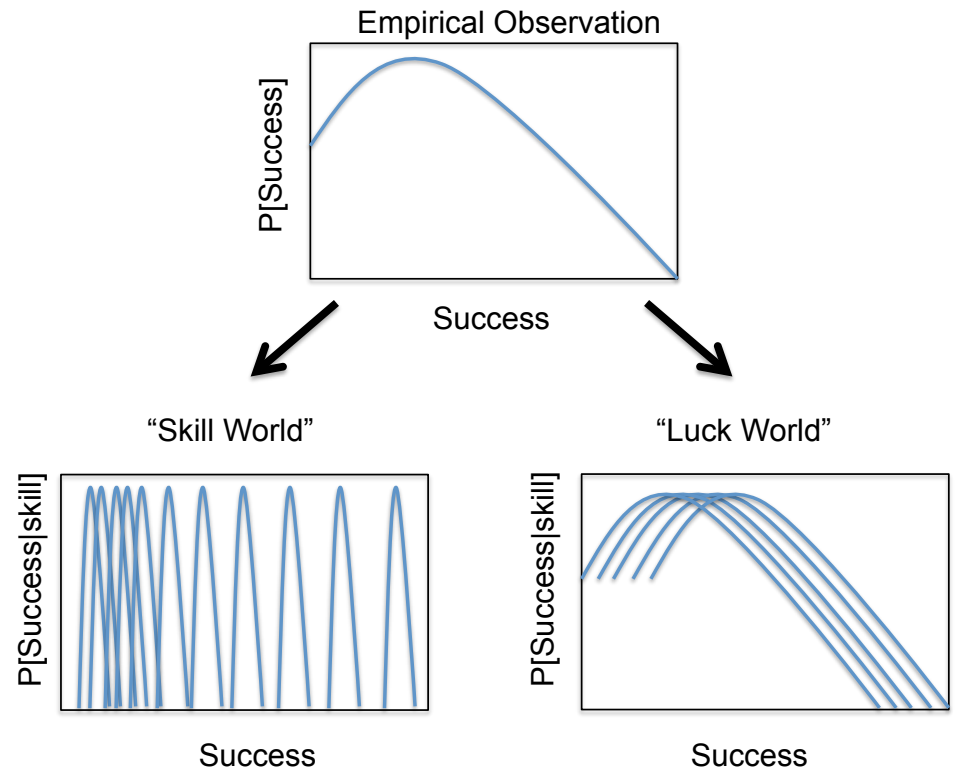
A unified framework: Luck vs. skill[2]

- Model success S as a mix of skill Q and luck ϵ:

$$S = f(Q) + \epsilon$$

- Measure the fraction of variance remaining after conditioning on skill:

$$F = \frac{\mathbb{E}[\mathrm{Var}(S|Q)]}{\mathrm{Var(S)}} = 1 - R^2$$

- $R^2 = 1$ in a pure skill world, $R^2 = 0$ in pure luck world

Empirical Observation

"Skill World"   "Luck World"

P[Success] vs Success

P[Success|skill] vs Success

---

[2]Formalizes Maboussin (2012)

Data

- Examined all 1.4B tweets containing URLs posted in February 2015

- Eliminated spam using internal Microsoft classifier

- Restricted attention to tweets containing URLs from the top 100 English-speaking domains with the most unique adopters

- Resulted in 850M tweets from 50M distinct users covering news, entertainment, videos, images, and products

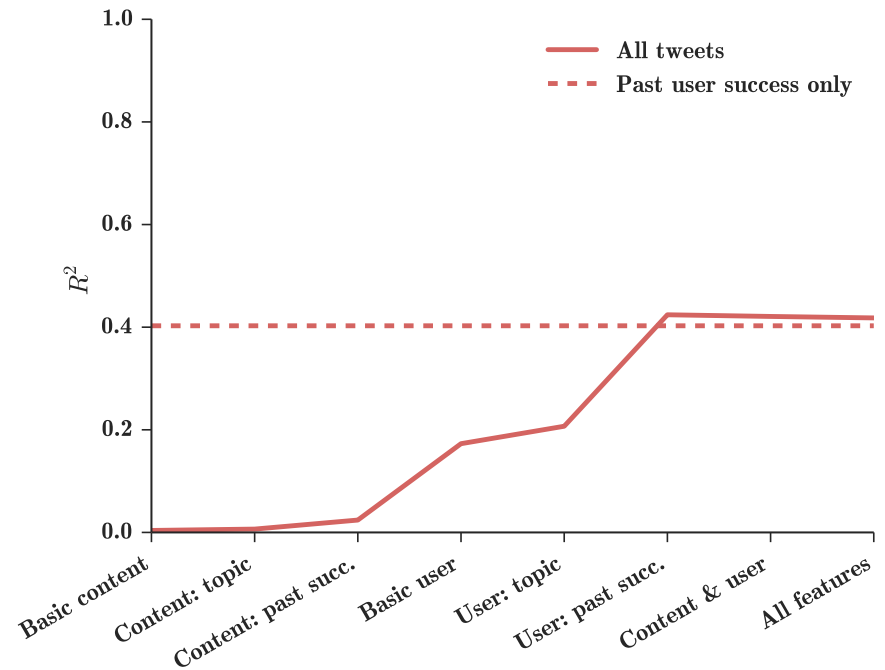- Measured the total cascade size for each seed tweet

Predictive features

Used a random forest to estimate success (cascade size)
given skill (available features)

- Basic content features: URL domain, time of tweet, spam score, ODP category

- Basic user features: number of followers, number of friends, number of posts, account creation time

- Topic features: the most probable Latent Dirichlet Allocation topic for each user and tweet, along with an interaction term

- Past success: the average number of retweets received by each URL and user in the past

Predictive performance

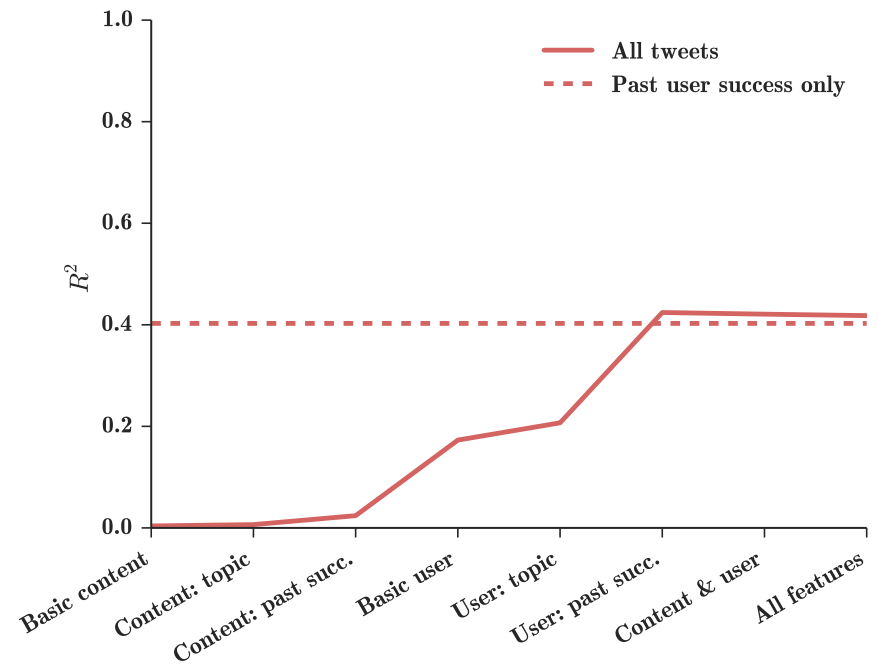best model explains roughly half of the variance in outcomes

Predictive performance

Content features alone perform poorly

Model	Tweet time	Domain	Spam score	Category	Tweet topic	Past url success	User time	Followers	Friends	Statuses	User topic	Past user success	Topic interaction
1. Basic content	✓	✓	✓	✓									
2. Content, topic	✓	✓	✓	✓	✓								
3. Content, past succ.	✓	✓	✓	✓	✓	✓							
4. Basic user							✓	✓	✓	✓			
5. User, topic							✓	✓	✓	✓	✓		
6. User, past succ.							✓	✓	✓	✓	✓	✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Predictive performance

Basic user features provide a reasonable boost in performance
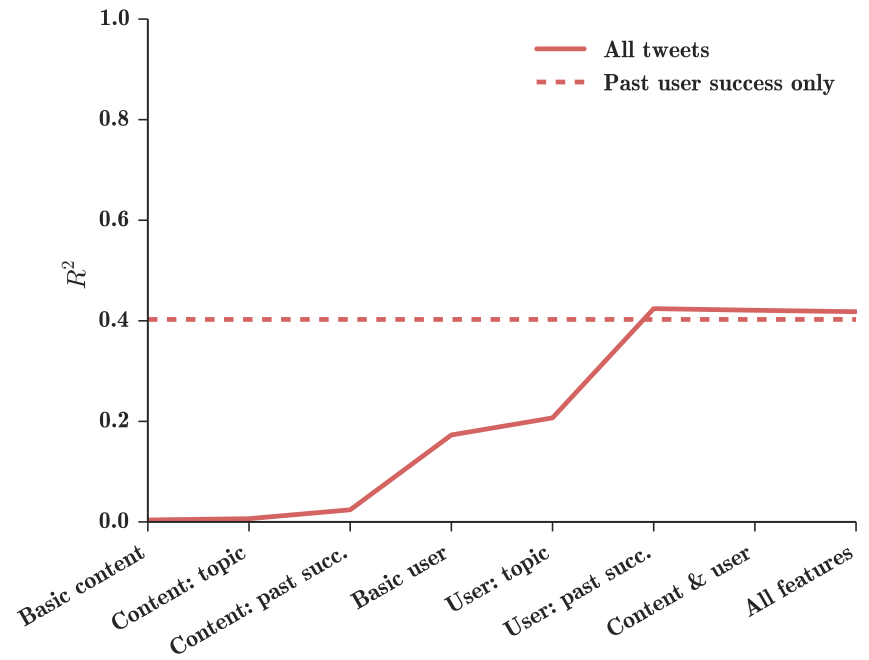
Model	Tweet time	Domain	Spam score	Category	Tweet topic	Past url success	User time	Followers	Friends	Statuses	User topic	Past user success	Topic interaction
1. Basic content	✓	✓	✓	✓									
2. Content, topic	✓	✓	✓	✓	✓								
3. Content, past succ.	✓	✓	✓	✓	✓	✓							
4. Basic user							✓	✓	✓	✓			
5. User, topic							✓	✓	✓	✓	✓		
6. User, past succ.							✓	✓	✓	✓		✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Predictive performance

Past user success alone accounts for almost all of predictive power
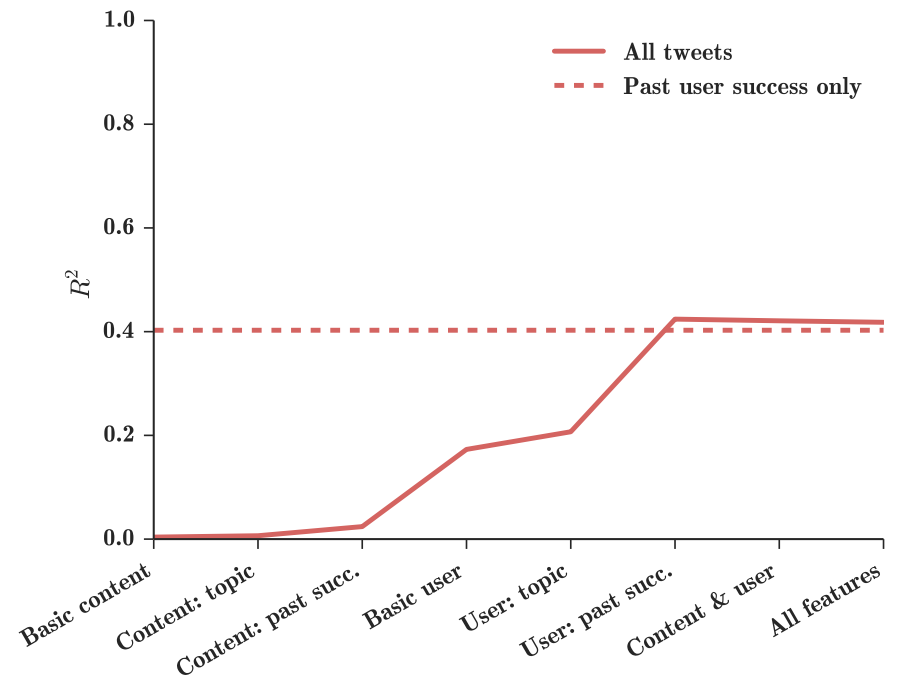
Model	Tweet time	Domain	Spam score	Category	Tweet topic	Past url success	User time	Followers	Friends	Statuses	User topic	Past user success	Topic interaction
1. Basic content	✓	✓	✓	✓									
2. Content, topic	✓	✓	✓	✓	✓								
3. Content, past succ.	✓	✓	✓	✓	✓	✓							
4. Basic user							✓	✓	✓	✓			
5. User, topic							✓	✓	✓	✓	✓		
6. User, past succ.							✓	✓	✓	✓	✓	✓	
7. Content, user	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
8. All	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

- Both models derive their predictive power from the same simple feature: a user's past success

- Content features are only weakly informative

- Performance plateaus as we add more features, suggesting a possible limit to the predictability of diffusion outcomes

How can you *prove* a limit?

- Results robust to other ML models
  - Decision tree, linear regression
- Consistent with prior work
- Asymptote, dependency between features
- Can't rule everything out
  - Simulation

Simulation

- SIR disease model
- Scale free network similar to Twitter
  - 7M users, α = 2.05
  - 8B simulated cascades
- *Quality*: R_0 = average neighbors infected
  - p(infect over edge) x *mean-degree*
- Prediction task
  - Given (possibly noisy) estimate of R_0 and the seed node, predict cascade size

Conclusions

Most things don't spread, but when they do, it's difficult to predict success

Conclusions

Despite a great deal of research on the topic, it's difficult to assess long-term progress in predicting success

Conclusions

State-of-the-art models explain roughly half of the variance in outcomes, based primarily on past success

Conclusions

This is likely due to randomness in diffusion process itself, rather than our ability to estimate or model it

Detecting influenza epidemics using search engine query data

Jeremy Ginsberg, Matthew H. Mohebbi ✉, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski & Larry Brilliant

> ⓘ This article has been updated

Abstract

Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists

Keyword, Author, or DO

Home   **Articles**   **Front Matter**   **News**   **Podcasts**   **Authors**   Sub

RESEARCH ARTICLE

Accurate estimation of influenza epidemics using Google search data via ARGO

Shihao Yang, Mauricio Santillana, and S. C. Kou

+ See all authors and affiliations

| Article | Figures & SI | Info & Metrics | | 🗋 PDF |

Significance

Big data generated from the Internet have great potential in tracking and predicting massive social activities. In this article, we focus on tracking influenza epidemics. We

RESEARCH ARTICLE

Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales

Donald R. Olson ✉, Kevin J. Konty, Marc Paladini, Cecile Viboud, Lone Simonsen

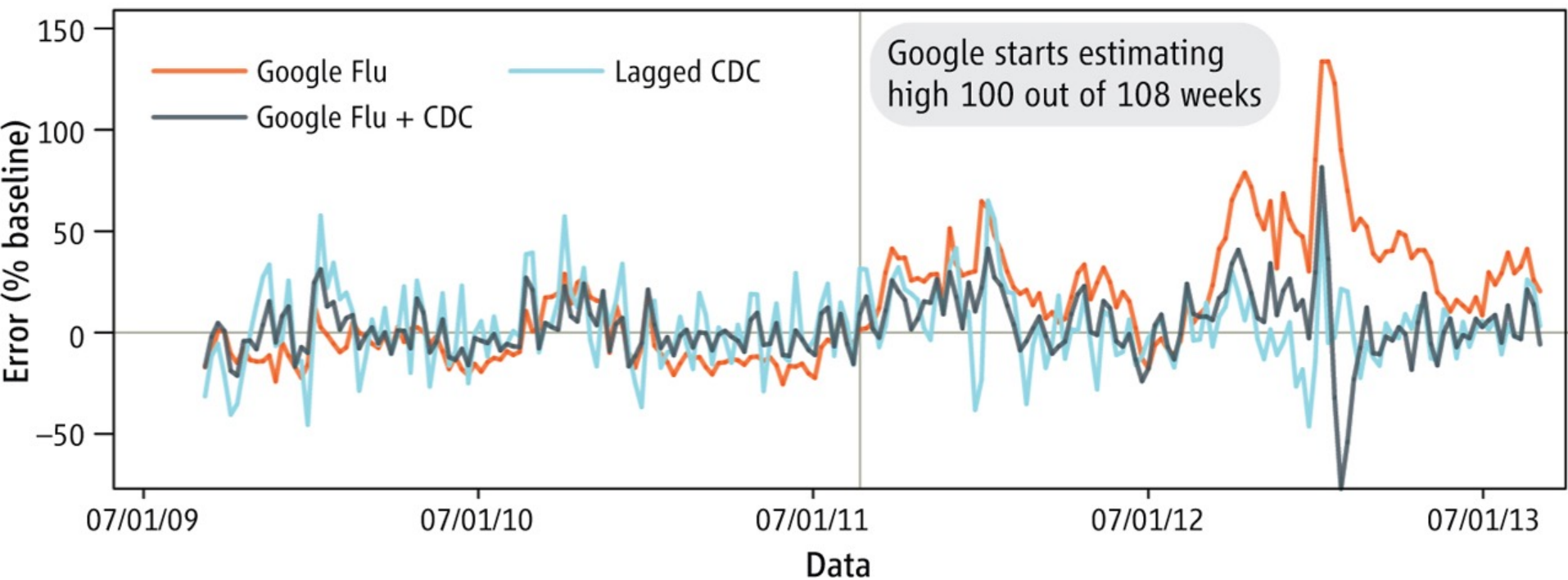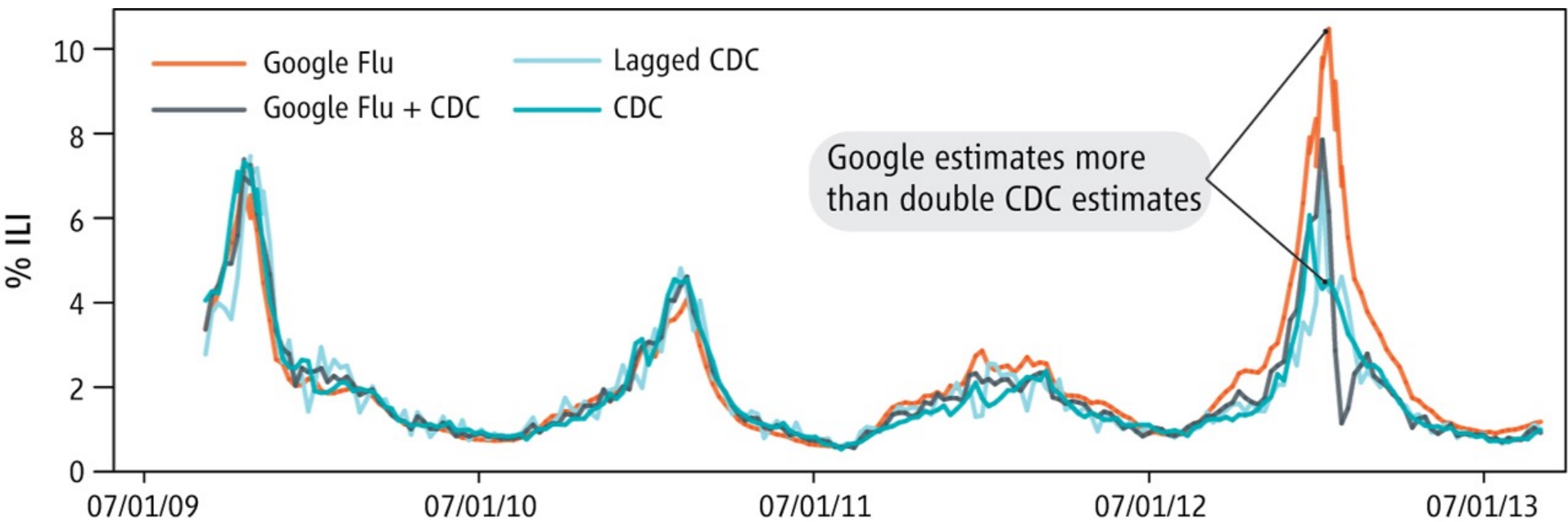| Article	Authors	Metrics	Comments	Media Coverage

Abstract

The goal of influenza-like illness (ILI) surveillance is to determine the timing, location and magnitude of outbreaks by monitoring the frequency and progression of clinical case incidence. Advances in computational and information technology have allowed for automated collection of higher volumes of electronic data and more timely analyses than previously possible. Novel surveillance systems, including those based on internet search query data like Google Flu Trends (GFT), are being used as surrogates for clinically-based reporting of influenza-like-illness (ILI). We investigated the reliability of GFT during the last decade (2003 to 2013), and compared weekly public health surveillance with search query data to characterize the timing and intensity of seasonal and pandemic influenza at the national (United States), regional (Mid-Atlantic) and local (New York City) levels. We identified substantial flaws in the original and updated GFT models at all three geographic scales, including completely missing the first wave of the 2009 influenza A/H1N1 pandemic, and greatly overestimating the intensity of the A/H3N2 epidemic during the 2012/2013 season. These results were obtained for both the original (2008) and the updated (2009) GFT algorithms. The performance of both models was

Figure: Weekly influenza-like illness (ILI) percent plotted over influenza seasons from 2003/2004 through 2012/2013 for three panels: A — United States, B — Mid-Atlantic States, C — New York City.

Legend:
- Sentinel and syndromic surveillance ILI-percent
- Serfling model ILI-percent baseline and 95% threshold
- Original GFT model-fitting and prospective periods
- Updated GFT model-fitting and prospective periods

X-axis: Date (influenza season)
Y-axis: Weekly influenza-like illness (ILI) percent

The parable of google flu

Top panel: Legend — Google Flu, Google Flu + CDC, Lagged CDC, CDC. Y-axis: % ILI (0 to 10). X-axis: Data (07/01/09 to 07/01/13). Annotation: "Google estimates more than double CDC estimates"

Bottom panel: Legend — Google Flu, Google Flu + CDC, Lagged CDC. Y-axis: Error (% baseline) (−50 to 150). X-axis: Data (07/01/09 to 07/01/13). Annotation: "Google starts estimating high 100 out of 108 weeks"

Big data hubris

Algorithmic Dynamics

It's Not Just About Size of the Data

danah boyd & Kate Crawford

CRITICAL QUESTIONS FOR BIG DATA
Provocations for a cultural, technological, and scholarly phenomenon

The era of Big Data has begun. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions. Diverse groups argue about the potential benefits and costs of analyzing genetic sequences, social media interactions, health records, phone logs, government records, and other digital traces left by people. Significant questions emerge. Will large-scale search data help us create better tools, services, and public goods? Or will it usher in a new wave of privacy incursions and invasive marketing? Will data analytics help us understand online communities and political movements? Or will it be used to track protesters and suppress speech? Will it transform how we study human communi-

Journal Home | Browse Journal ⌄ | Journal Info ⌄ | Stay Connected ⌄ | **Submit Paper** | Search 🔍

Deeper data: a response to boyd and Crawford

Andre Brock

Article information ⌄

Altmetric 5 🔒

Abstract

Data analysis of any sort is most effective when researchers first take account of the complex ideological processes underlying data's originating impetus, selection bias, and semiotic affordances of the information and communication technologies (ICTs) under examination.

Keywords

Big Data, critical cultural informatics, critical information studies, data and society, digital sociology, social media and society

In 2013, Lois Scheidt and I organized a panel for the International Congress of Qualitative Inquiry titled 'Small data in a big data world' as a response to 'Six Provocations for Big Data'. Our panelists presented incredible work conceptualizing new approaches in an age of 'big data' to qualitative social media research,

ESSAY

Prediction and explanation in social systems

Jake M. Hofman,* Amit Sharma,* Duncan J. Watts*

Historically, social scientists have sought out explanations of human and social phenomena that provide interpretable causal mechanisms, while often ignoring their predictive accuracy. We argue that the increasingly computational nature of social science is beginning to reverse this traditional bias against prediction; however, it has also highlighted three important issues that require resolution. First, current practices for evaluating predictions must be better standardized. Second, theoretical limits to predictive accuracy in complex social systems must be better characterized, thereby setting expectations for what can be predicted or explained. Third, predictive accuracy and interpretability must be recognized as complements, not substitutes, when evaluating explanations. Resolving these three issues will lead to better, more replicable, and more useful social science.

For centuries, prediction has been considered an indispensable element of the scientific method. Theories are evaluated on the basis of their ability to make falsifiable predictions about future observations—observations that come either from the world at large or from experiments designed specifically to test the theory. Historically, this process of prediction-driven explanation has proven uncontroversial in the physical sciences, especially in cases where theories make relatively unambiguous predictions and data are plentiful. Social scientists, in contrast, have generally deemphasized the importance of prediction relative to terms of predictive accuracy. We believe that the confluence of these two trends presents an opportune moment to revisit the historical separation of explanation and prediction in the social sciences, with productive lessons for both points of view. On the one hand, social scientists could benefit by paying more attention to predictive accuracy as a measure of explanatory power; on the other hand, computer scientists could benefit by paying more attention to the substantive relevance of their predictions, rather than to predictive accuracy alone.

Standards for prediction

contained links to the top 100 most popular websites, as measured by unique visitors. In addition to holding the data set fixed, for simplicity, we also restricted our analysis to a single choice of model, reported in (*11*), that predicts cascade size as a linear function of the average past performance of the "seed" individual (i.e., the one who initiated the cascade). Even with the data source and model held fixed, Fig. 1 (top) shows that many potential research designs remain: Each node represents a decision that a researcher must make, and each distinct path from the root of the tree to a terminal leaf node represents a potential study (*12*). We emphasize that none of these designs is intrinsically wrong. Nevertheless, Fig. 1 (bottom) shows that different researchers—each making individually defensible choices—can arrive at qualitatively different answers to the same question. For example, a researcher who chose to measure the AUC [the area under the receiver operating characteristic (ROC) curve] on a subset of the data could easily reach the conclusion that their predictions were "extremely accurate" [e.g., (*10*)], whereas a different researcher who decided to measure the coefficient of determination (R^2) on the whole data set would conclude that 60% of variance could not be explained [e.g., (*6*)].

Reality is even more complicated than our simple example would suggest, for at least three reasons. First, researchers typically start with different data sets and choose among potentially many different model classes; thus, the schematic in Fig. 1 is only a portion of the full design space. Second, researchers often reuse the same data set to assess the out-of-sample

Class Exercise

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., the latest version of iPhone), as observed on social media
ii) Whether greater anonymity leads to greater hate speech on social media
iii) Whether following recommended videos on YouTube leads down a more politically extreme rabbit hole
iv) Whether deplatforming reduces misinformation on social media

Onward…

How do we decide what should and shouldn't be predicted?

Private Traits and Attributes are Predictable from Digital Records of Human Behavior
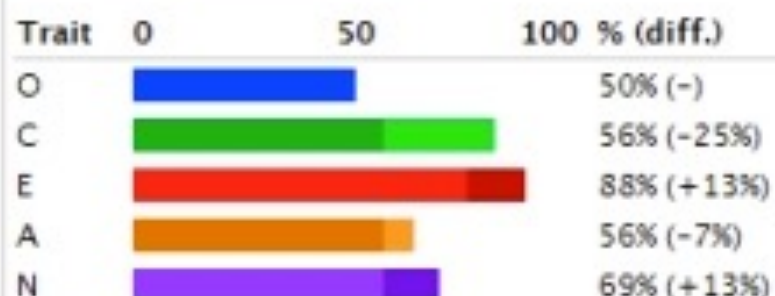
Your Friends` Personalities

Most Like Me

Your Personality Soulmate

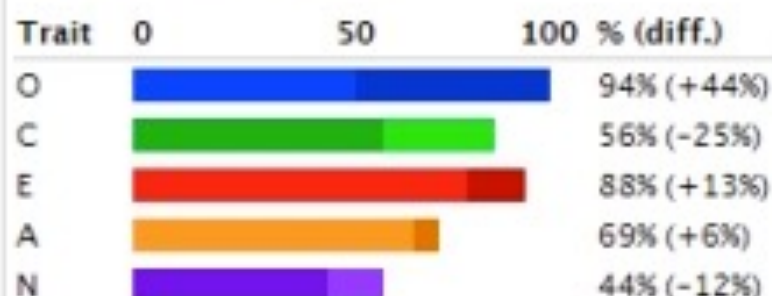Sofie Jansson

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

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

Least Like Me

Maybe Opposites Attract?

Damon Alexander Young

Similarity Score: **75.87%**
(How was this calculated?)

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

Friend`s Name	Personality					Similarity Score	View Comparison Graph	View Full Personality Profile	Friend Rating Status
	O	C	E	A	N				
You	50%	81%	75%	63%	56%				
Sofie Jansson	50%	56%	88%	56%	69%	86%			
Sara Lee	88%	63%	63%	69%	56%	80%			
Damon Alexander Young	94%	56%	88%	69%	44%	76%			

The Plus-es 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?

Who benefits from this research?

BBC

News   Sport   Weather   Capital   Future   Shop   TV   Radio   More…

Search

NEWS TECHNOLOGY

Home | US & Canada | Latin America | UK | Africa | Asia | Australia | Europe | Mid-East | Business | Health | Sci/Environment | Tech | Entertainment | Video

11 March 2013 Last updated at 15:00 ET

Share f 🐦 🖨

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

Facebook changes privacy settings

Privacy groups lobby Facebook

BBC

News Sport Weather Capital Future Shop TV Radio More... Search

NEWS TECHNOLOGY

Home | US & Canada | Latin America | UK | Africa | Asia | Australia | Europe | Mid-East | Business | Health | Sci/Environment | Tech | Entertainment | Video

"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."

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

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

💬 Comment

sociological science
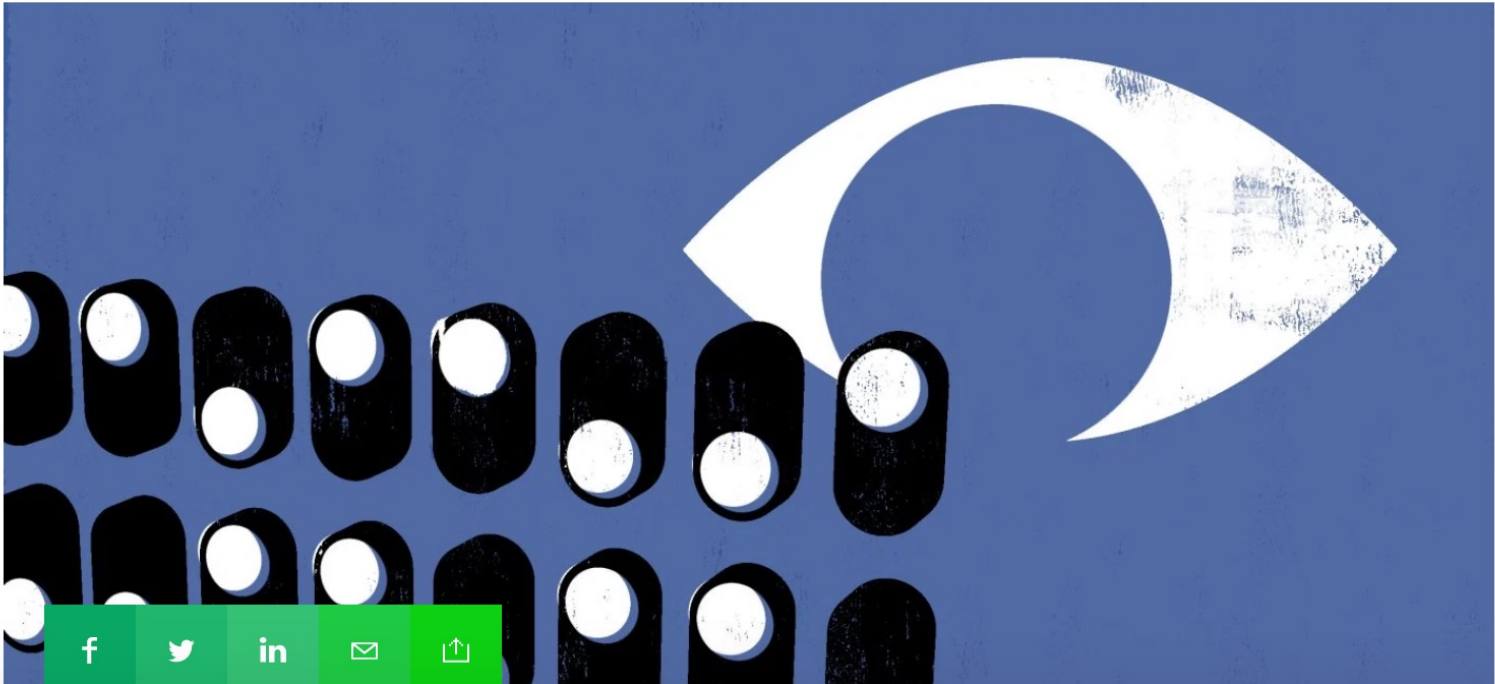
Gaydar and the Fallacy of Decontextualized Measurement

Andrew Gelman,[a] Greggor Mattson,[b] Daniel Simpson[c]

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 we 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 measure of perception of sexual orientation. Instead, we argue that the concept of gaydar inherently exists 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

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

Page: 1 of 47 — Automatic Zoom

DEEP NEURAL NETWORKS CAN DETECT SEXUAL ORIENTATION FROM FACES

1. THIS IS A PREPRINT OF THE PEER REVIEWED ARTICLE TO APPEAR IN JOURNAL OF
2. PERSONALITY AND SOCIAL PSYCHOLOGY.
3.
4. THE MOST RECENT VERSION IS AVAILABLE AT https://osf.io/zn79k/
5. AUTHOR NOTES ARE AVAILABLE AT: https://goo.gl/9b2aR2
6.
7. Deep neural networks are more accurate than humans at detecting sexual orientation from facial
8. images
9.
10. Yilun Wang, Michal Kosinski
11. Graduate School of Business, Stanford University, Stanford, CA94305, USA
12. michalk@stanford.edu
13.
14.
15. *The study has been approved by the IRB at Stanford University*
16.
17. Citation: Wang, Y., & Kosinski, M. (in press). Deep neural networks are more accurate than
18. humans at detecting sexual orientation from facial images. *Journal of Personality and*

wang_kosinski.pdf

Download preprint — Version: 10

Abstract

We show that faces contain much more information about sexual orientation than can be perceived and interpreted by the human brain. We used deep neural networks to extract features from 35,326 facial images. These features were entered into a logistic regression aimed at classifying sexual orientation. Given a single facial image, a classifier ...

See more

Preprint DOI

10.17605/OSF.IO/HV28A

License

No license ▸

Disciplines

Social and Behavioral Sciences   Psychology

Tags

Artificial Intelligence   Big Data   Computational Social Science   Computer Vision   Facial recognition   Prenatal Hormone Theory   Privacy   Sexual orientation

Citations

APA

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

Automatic Crime Prediction using Events Extracted from Twitter Posts

Xiaofeng Wang, Matthew S. Gerber, and Donald E. Brown

Department of Systems and Information Engineering, University of Virginia
{xw4u,msg8u,brown}@virginia.edu

Abstract. Prior work on criminal incident prediction has relied primarily on the historical crime record and various geospatial and demographic information sources. Although promising, these models do not take into account the rich and rapidly expanding social media context that surrounds incidents of interest. This paper presents a preliminary investigation of Twitter-based criminal incident prediction. Our approach is based on the automatic semantic analysis and understanding of natural language Twitter posts, combined with dimensionality reduction via latent Dirichlet allocation and prediction via linear modeling. We tested our model on the task of predicting future hit-and-run crimes. Evaluation results indicate that the model comfortably outperforms a baseline model that predicts hit-and-run incidents uniformly across all days.

1 Introduction

Traditional crime prediction systems (e.g., the one described by Wang and Brown [14]) make extensive use of historical incident patterns as well as layers of in-

Onward...

How do we decide what should and shouldn't be predicted?

-- Importance of considering ethics and privacy in social computing