Building Socio-technical Systems for Social Good

Jane Im

September 14, 2019

For my PhD I aim to **design and build novel social computing systems for social good**. To be specific, I aspire to build new socio-technical systems that can **protect and empower vulnerable population** (women, people of color, LGBTQ community). Below I outline the major components of my past and current work as well as how I aim to move forward.

1. Past work

My past work helped me to solidify my research interests in resolving problems in online communities and social media platforms (1.1, 1.2) and also resulted into a first author paper accepted to CSCW. I have also gained expertise in building systems that led me to have interest in system-building aspect of HCI (1.3).

1.1 Improving dispute resolution on Wikipedia

Resolving disputes are crucial to large-scale collaborative work in online communities including Wikipedia. Prior to starting my PhD, I worked on improving a formal dispute resolution process on Wikipedia which resulted in a first-author publication at **CSCW 2018** [1]. I investigated the challenges Wikipedia editors face when trying to resolve disputes through Requests for Comments (RfC), a common process used by Wikipedia editors for requesting input from uninvolved editors concerning disputes about policy, guideline, or article content. I used mixed-methods of gathering and quantitatively analyzing an exhaustive dataset of RfCs on English Wikipedia and qualitatively analyzing a subset of threads of unresolved disputes, as well as helping conduct interviews with Wikipedia editors actively involved in the process of RfC[2].

Results showed that a third of RfCs remain stale (meaning disputes are unresolved), with various reasons spanning from participants' behavioral problems such as bickering to the RfC thread becoming too contentious and complicated. Deriving features from the reasons discovered, I further built machine learning models that predict whether a given RfC will reach consensus within time. The models can **inform editors how likely the RfC is going to successfully end as time passes by, along with the features that will help in boosting the likelihood**. When presenting this work in Wikimedia

[1] Im, J., Zhang, A. X., Schilling, C. J., and Karger, D. (2018). Deliberation and resolution on wikipedia: A case study of requests for comments. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):74

[2] Amy Zhang, a co-author of the paper, lead the interviews.

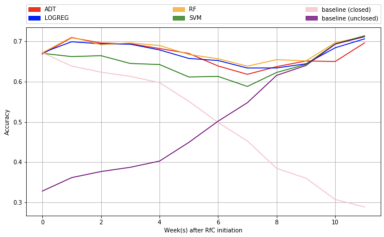

Figure 1: I developed a model that predicts whether an RfC will go stale with 75.3% accuracy, a level that is approached as early as one week after dispute initiation.


RFC on income inequality effects

*The following discussion is closed. **Please do not modify it.** Subsequent comments should be made on the appropriate discussion page. No further edits should be made to this discussion.*

Consensus was to "omit" the material due to concerns that it is off topic. Morphh (talk) *14:31, 17 January 2014 (UTC)*

Instead of edit-warring over this excerpt, we clearly need an RFC.

Inequality in land and income ownership is negatively correlated with subsequent economic growth. A strong demand for redistribution will occur in societies where a large section of the population does not have access to the productive resources of the economy. Rational voters must internalize such issues. (Alesina, Alberto (1994). "Distributive Politics and Economic Growth" (PDF). *Quarterly Journal of Economics*. **109** (2): 465–90. doi:10.2307/2118470. Retrieved 17 October 2013. Unknown parameter |coauthors= ignored (|author= suggested) (help); Unknown parameter |month= ignored (help)) High unemployment rates have a significant negative effect when interacting with increases in inequality. Increasing inequality harms growth in countries with high levels of urbanization. High and persistent unemployment also has a negative effect on subsequent long-run economic growth. Unemployment may seriously harm growth because it is a waste of resources, because it generates redistributive pressures and distortions, because it depreciates existing human capital and deters its accumulation, because it drives people to poverty, because it results in liquidity constraints that limit labor mobility, and because it erodes individual self-esteem and promotes social dislocation, unrest and conflict. Policies to control unemployment and reduce its inequality-associated effects can strengthen long-run growth. (Castells-Quintana, David (2012). "Unemployment and long-run economic growth: The role of income inequality and urbanisation" (PDF). *Investigaciones Regionales*. **12** (24): 153–173. Retrieved 17 October 2013. Unknown parameter |coauthors= ignored (|author= suggested) (help))

Should that be included in the Economic effects/Income inequality section? EllenCT (talk) 02:25, 2 January 2014 (UTC)

Survey

- **Support** inclusion of the passage as a separate paragraph, to explain why income equality is a positive economic effect. EllenCT (talk) 02:25, 2 January 2014 (UTC)
- **Omit** the paragraph in its entirety, as it does not even approach the subject of taxation, progressive or otherwise. Obviously off-topic and superfluous. Roccodrift (talk) 02:29, 2 January 2014 (UTC)

Figure 2: Example of a resolved dispute through Request for Comment (RfC).

Research Showcase[3], Wikipedia editors showed interest in using an improved version of models in Wikipedia's system so the editors can use it while participating in RfCs. For instance, if the average account age of participants has the highest feature importance currently, this indicates that experienced editors should be invited to participate in the RfC in order to improve the likelihood of it reaching consensus.

Through this project I realized I am broadly interested in building computational tools to help resolving problems in online communities. Furthermore, the process of publishing a paper at a top-tier conference solidified my interest in research.

[3] https://www.mediawiki.org/wiki/Wikimedia_Research/Showcase#October_2018

1.2 Detecting Russian troll accounts on Twitter

During my first semester of PhD, I had the opportunity to focus on an impactful project of combating state-sponsored propaganda on social platforms [4]. U.S. intelligence agencies, U.S. courts, and researchers have found that Russia's Internet Research Agency (IRA) tried to interfere with the 2016 U.S. election by running fake accounts—often called the "Russian troll" accounts—on various social media platforms including Twitter [5]. Since one of the first steps to combat such accounts is to find them, I built machine learning models that can identify Russian trolls on Twitter using an unbalanced dataset of 2.2K Russian troll accounts released by Twitter and 170K control accounts. Using over 5000 features which include behavioral and linguistic features, our classifiers achieve an AUC of 98% and precision of 78%. I believe the models I built is a proof-of-concept in showing that there is a way to fight against state-sponsored propaganda by easily identifying Russian trolls on social platforms, especially on a large-scale.

[4] Im, J., Chandrasekharan, E., Sargent, J., Lighthammer, P., Denby, T., Bhargava, A., Hemphill, L., Jurgens, D., and Gilbert, E. (2019). Still out there: Modeling and identifying russian troll accounts on twitter. *arXiv preprint arXiv:1901.11162*

[5] Gorodnichenko, Y., Pham, T., and Talavera, O. (2018). Social media, sentiment and public opinions: Evidence from# brexit and# uselection. Technical report, National Bureau of Economic Research

1.3 Extending MIT App Inventor to Support Modular Code

Prior to starting my PhD, I also sought to enable novice programmers to create modular code in the MIT App Inventor under the guidance of Hal Abelson. MIT App Inventor[6] is a web platform where millions of users build Android Apps using blocks-based programming. This can significantly assist novice programmers in creating diverse apps. However, the limited number of blocks usually prevents novice programmers from building specific goal-oriented apps that require blocks that currently do not exist in the system. I developed **customized blocks** in the system so that users can freely import any JavaScript API for each project, and then design and add customized blocks that execute the API's functions. Especially regarding that the end-users will be novice programmers, I extended the system so that users could intuitively build a customized block by snapping lower-level blocks that form it.

Through this work, I realized I care deeply about designing and developing large-scale systems that are easily accessible to a wide range of users. This further solidified my interest in studying the system-building aspect of HCI.

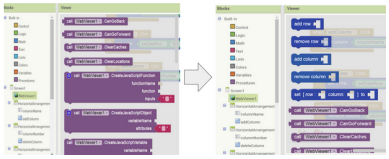

Figure 3: Customized blocks enable novice programmers to build a wide range of apps using MIT App Inventor.

[6] `https://appinventor.mit.edu/`

2. Current Work

While my previous research experience helped me to discover my broad interests and hone skills for building computational tools and systems, my ongoing work helped me narrow down my passion to building systems that can **protect and empower vulnerable population including women, LGBTQ community, and people of color**.

2.1 Deep Social Signals

We rely on signals when interacting with "strangers" on social platforms. One main area where we can earn important information about an account is from each account's history of posts. However, it takes a high receiver cost [7] for a person to manually digest an account's history of posts. To address this challenge I propose a new social signal called "**deep social signals (DSS)**", which are **social signals computed from a user's history of posts using algorithms**. Unlike conventional signals such as self-descriptions in profiles [8], deep social signals aim to i) reduce receiver costs for gathering information from an account's history of posts and ii) increase the cost for faking information.

To demonstrate the concept, we built Sig, a Chrome extension that computes and visualizes deep social signals. We conducted a field study of Sig on Twitter which surfaced DSS of toxicity and misinfor-

[7] Guilford, T. and Dawkins, M. S. (1991). Receiver psychology and the evolution of animal signals. *Animal behaviour*, 42(1):1–14

[8] Donath, J. (2007). Signals in social supernets. *Journal of Computer-Mediated Communication*, 13(1):231–251

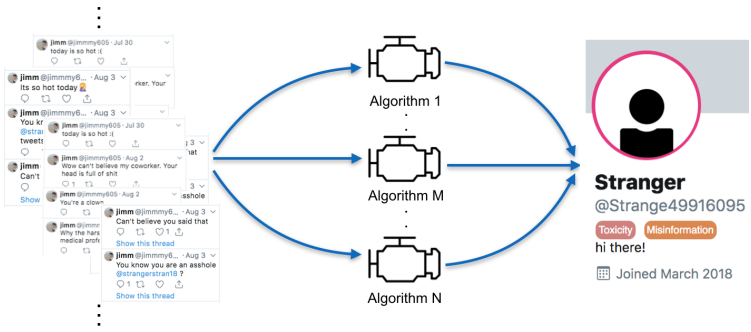

Figure 4: Concept of deep social signals (DSS).

mation. The results showed participants finding Sig overall useful in interacting with stranger accounts on Twitter. Many participants thought Sig provided reliable information of accounts in advance and some reported the extension helped them feel safe against toxic or misinformation-spreading accounts. Some participants even actively used Sig to follow (or decide not to follow), mute or even block toxic or misinformation-spreading accounts during the field study.

Furthermore, many field study participants expressed that **DSS of toxicity such as harassing behavior would be especially helpful to vulnerable population including women, LGBTQ community, and people of color**. Previous research has already shown that people of color, women, or LGBT communities suffer most from online harassment [9]. The field study participants' responses motivated me to think about proposing new concepts and building systems that can help vulnerable population. I am currently in the process of submitting this work to **CHI 2020**.

Figure 5: Sig flags an account if at least one deep social signal indicates it might be risky to interact with the account. For the field study we tested Sig supporting DSS of toxicity and misinformation.

[9] Duggan, M. (2017). Online harassment 2017

2.2 *Embedding Interpersonal Consent into Social Platforms*

My pre-candidacy research focuses on embedding interpersonal consent into social platforms by taking a feminist point of view. That is, I aim to **design and build a social platform where interpersonal consent is at the core**, instead of trying to patch up consent-lacking ones behindhand. The motivation behind this research is the hope of protecting vulnerable users from online harassment and abuse by protecting their consent, as well as empowering them to fully enjoy social platforms (instead of abandon using them due to fear of harassment).

So far, I have been working on **defining what interpersonal consent means in the context of social platforms** based on previous literature of feminism, criminology, and CSCW. Up to now, I have defined the core concepts of consent as **voluntary**, **informed**, **unburdensome**, **ongoing (reversible)**, **specific**, and **competence**. During

the process I have submitted a workshop paper to **CSCW 2019** [10] and talked with other scholars on how to conduct research with vulnerable population and sensitive topic.

I'm also working on building a generative framework – a set of design insights derived from the concepts of consent – that can be used in building social platforms which are consent-embedded from the start. Based on the generative framework, I aim to build a prototype of a social platform where consent is embedded in every layer of the software.

[10] Im, J. (2018). Non-consensual images & videos and consent in social media. *CSCW 2018 workshop*