Thesis proposals

Ideas for research from social bots to beer

Florian Daniel
florian.daniel@polimi.it

February 25, 2019
Detection of harmful social bots

Identification of harmful communication patterns

Conversational screen readers

Domain-specific content extraction
**Social** Bot = algorithmically driven entity that behaves like a human in online communications
Empirical study shows: bots may cause harm to humans

What bots do

Chat
- Talk with user
- Redirect user

Post
- Write post
- Comment post
- Forward post

Endorse
- Like message
- Follow user

Participate
- Create user

Action

Examples
- eCommerce, CustomerSvc
- BoostJuice
- AASlang
- SMSsex
- Juice
- Oreo
- SethRich
- DeathThreat
- Death

When abuse happens

- Disclose sensitive facts
- Denigrate
- Be grossly offensive
- Be indecent or obscene
- Be threatening
- Make false allegations
- Deceive
- Spam
- Spread misinformation
- Mimic interest
- Clone profile
- Invite space

What else can they do?

- Invade space

What else can go wrong?

- Not regulated

Regulated by law

References:
Detection of harmful social bots

How do we identify and classify bots according to the harm they may cause?
What has been done so far?
News-Spreader

- Propagandistic profile picture and tile
- High number of interactions
- Propaganda hashtags and messages in description
- Lots of (political) news retweeting

Spam-Bot

- Profile picture with corporate logo
- Profile tile with catchy messages
- Corporate link in description paired with messages of job offers (product sales)
- High number of tweets with high frequency
- Repetitive tweets aimed at spamming URLs

Fake-Follower

Optional profile picture (usually missing)

Low number of content posting (usually null)
Following as main interaction

Optional (re)tweeting activity

Poor description (usually empty)

Poor profile customization

Thesis ingredients

Creation of datasets

Feature selection and engineering

Multi-class ensemble classifier

Figure 7.3: Client - Server architecture

Inception neural network. Users with less than 10 tweets, or less than 10 tweets with embedded images, require a shorter timespan to be classified, which is, in average up to 5 seconds. Figure 7.4 shows the repartition of computational time, along with the processes involved. It is based on the timespan needed to classify users with at least 100 tweets and with at least 10 tweets with images.

7.2.1 Engine

The engine of the web application is a Python 3 script. It performs all the steps described in the pipeline execution section 6.2. The models, that have been previously fitted with data, serialized and stored, are now loaded by the Python script. They have to perform a single prediction at a time. In addition to the models we built for the classification, the pre-trained convolutional neural network for NSFW recognition has been introduced to the pipeline, in order to infer on the media contents posted by the examined user. The first step consists in calling the Twitter APIs to retrieve user's data and its most recent tweets, up to 100. The script, then, handles the

Can we add more types of harm? How to train classes as users use BotBuster?
Which potential harms can we identify inside the code of the bots?
Harmful code patterns

Twitter

Python

Next: patterns search engine + web site for users

What else can we learn from the code of bots? Is it possible to trace back from messages to code?

Can we extract conversational knowledge from webpages?
What about “talking” to websites?
Domain-specific content extraction

How effectively can we extract recipes from free text? And how do we do it?
Typical data science steps

- Domain understanding
- Data collection
- Manual inspection of data
- Hypotheses formulation
- Feature engineering, data labeling
- Algorithm engineering (from AI/machine learning to statistics)
- Validation and hypotheses verification

Online tool
Thesis/project proposals

Are you looking for a project for your Master thesis or for some other project you would like to carry out during your studies, for example for the BSc course Progetto di Ingegneria Informatica? Below you can find some research topics I am or would like to work on. All these can be turned into projects and theses and are meant to allow you to get a glimpse behind the scenes of the work of a professor and to show you that there is more than teaching in a university. Methodologically, typically these topics are to be approached using a data science and/or web/software engineering methodology.

If you spot something you like, drop me an email or get in touch with me! I’ll be happy to discuss possible options to work together. Also, yes, you can work in groups of two and write and present your thesis together.

What you need: programming skills, knowledge of web technologies, data management skills, a brain and passion.

What you get: new competences, a thesis (or tesina), a project, personal satisfaction and recognition of your work.

Writing your thesis: yes, a thesis is also about writing and writing skills. Here some thoughts on writing and a possible template for your thesis.

Harmful social bots

Topics: social bots, social networks, data science, ethics

Bots are algorithmically driven entities that behave like humans in online communications. They are increasingly infiltrating social conversations on the Web or in chat apps. If not properly prevented, this presence of bots may cause harm to the humans they interact with, e.g., they may offend or discriminate people. The goal of this research is to understand which types of harm and which abuses may happen, whether abuses can be considered intentional or not, whether it is possible to prevent them and, if yes, how.

Research questions and possible thesis/project topics

- Bot detection: Given a specific type of abuse, e.g., discrimination, spamming or mimicking interest, how can we identify bots that may be vulnerable to this type of abuse? Answering this question requires following a data science methodology and may require the use of machine learning and AI algorithms, online/reinforcement learning techniques, network analysis techniques, and similar. Different social networks may be studied. The expected output are algorithms or methods able to classify social network accounts based on the harm they may cause.
- Code analysis: Here the idea is to start from the code of bots shared online, e.g., in GitHub, and to study which abusive code patterns can be identified, which effects they may have, and how to spot bots that implement them. Again, answering these questions requires a data science approach and may imply the