PROPOSTA DE CONSULTORIA: MOTOR PROJETO PEGABOT

Proponente

Grupo 1A - Desafio 1

Problema

Quais outras informações disponibilizadas pela API do Twitter podem agregar positivamente ao motor de análises do Pegabot?

Apresentação

O Pegabot possui um motor que utiliza informações disponibilizadas pela API do Twitter, de modo a identificar a probabilidade de uma conta de usuário ser ou não um bot. Percebemos que já há uma maneira hábil de classificar qual o tipo de bot possivelmente identificado, categorizando-o como malicioso ou não. Ainda assim, o motor do Pegabot pode avançar com aprimoramentos que permitem ainda maior acurácia ao detectar a probabilidade de uma conta ser ou não um bot.

As principais informações advindas da API do Twitter e que são utilizadas pelo Pegabot com o referido propósito são: tempo mediano entre tweets; retweets; número de amigos; tamanho do nome; menor tempo entre tweets; número de seguidores; origem da postagem; quantidade de hashtags; entre outros. Uma maneira de avançar, recorrendo a outros dados fornecidos pela API do Twitter seria a possibilidade de *analisar imagens e outros*

conteúdos de mídia, como figuras animadas (GIFs) e vídeos, que são compartilhados em tweets.

Nesse sentido, contas fakes consideradas maliciosas costumam espalhar imagens e vídeos com conteúdo que afeta negativamente os outros usuários, tirando vantagem do serviço de compartilhamento de mídia do Twitter, upando tais conteúdos com o suporte de hashtags e marcações de outros usuários para garantir a diluição de conteúdo malicioso na plataforma. (SAHOO e GUPTA, 2019). Não obstante, bots maliciosos também podem aproveitar-se de tal serviço com a finalidade de causar danos ao discurso digital, uma vez que imagens podem evocar emoções (NG e CARLEY, 2021) variadas de acordo com o intento do autor.

Portanto, analisar as informações de mídia contidas nos tweets, em conjunto com o compartilhamento de hashtags ou outras trends, pode ser relevante para identificar bots, aumentando o nível de precisão das probabilidades geradas pelo Pegabot. Tendo em vista que o uso de imagens e outras mídias pode aumentar o engajamento dos tweets (Wang *et al*., 2021) e que as emoções geradas através de arquivos de mídia podem acarretar na geração de violência *online* e *offline* (Carley, 2020), a ITS pode seguir um caminho parecido com o de Ng e Carley (2021). Estes elaboraram uma pesquisa para entender quais emoções eram passadas através de conteúdos de imagens e a diferença entre tais emoções transmitidas por contas legítimas e contas conhecidamente robotizadas.

Assim, propõe-se a realização de um teste de identificação de padrões nos conteúdos de mídia de contas identificadas como genuínas e como bots. Dentro da análise de conteúdo de mídias de contas consideradas como bots, classificá-los por tipo (malicioso ou não) e, então, reconhecer os padrões de compartilhamento de mídias de bots maliciosos.

Objetivo do desafio

Identificar os padrões de compartilhamento de mídias realizadas por bots no Twitter, para complementar o sistema de identificação do Pegabot e melhorar sua capacidade de detecção de bots maliciosos. Dessa forma, a capacidade de identificação de bots pode ser elevada, reconhecendo quais os tipos de temas de conteúdo de mídia compartilhados por bots do tipo maliciosos e quais as emoções dos usuários que podem surgir através dessas imagens.

Metodologia

O objetivo proposto pode seguir os seguintes passos para sua conclusão, tendo como base o trabalho realizado por Ng e Carley (2021):

1. Recolher dados da API do Twitter que estejam relacionadas a alguma trend específica da rede social, a qual reconhecidamente contém atuação de bots (ex.: Eleições 2022, Vacinas, Varíola dos macacos, Covid-19, etc.);
2. Para maior viabilidade de aplicação do método, é preferível que o estudo inicial seja concentrado na análise de apenas uma dessas trends/hashtags, além de se basear em um período específico (preferencialmente, aquele com maior volume de atividades e compartilhamento de mídias);
3. Identificar os tweets de bots e os de não-bots;
4. Identificar conjuntos de mídias que representem um tema;
5. Construir ou obter um classificador de emoções de mídias, a fim de categorizar tais dados por tipo de emoção (ex.: raiva, tristeza, alegria, surpresa, etc.);
6. Usando os conjuntos de temas e as categorias de emoções, analisar as diferenças entre as mídias compartilhadas entre os usuários legítimos e os bots, inclusive entre os bots considerados maliciosos e os não maliciosos;
7. Últimas observações: para ser um experimento confiável, é importante que um grande volume de dados seja analisado para detectar padrões.

Produto final

Um relatório é gerado ao fim do processo descrito acima, com os padrões de mídias reconhecidos e especificados, com as devidas sugestões de aprimoramento do Motor Pegabot.

Prazo de execução

Prazo indeterminado, pois depende da disponibilidade de equipe de programadores e do tempo padrão de execução das tarefas a serem realizadas.

Integrantes - GRUPO 1A:

Lêticia Lopes - Pesquisadora. Mestranda em Economia (UFPB). Bacharel em Finanças (UFC). Técnica em Contabilidade. Desenvolvendo pesquisa na área da Economia da Saúde e Políticas Públicas.

Lailson Viana - Pesquisador. Analista de dados Jr. (Itaú Unibanco). Mestrando em Economia (UFPB). Bacharel em Finanças (UFC) e em Administração (UNOPAR). Desenvolvendo pesquisa na área de Desigualdade de Renda.

Sophia Faro - Estagiária no setor de Jornalismo na Amazon Filmes. Graduanda em Comunicação Social - Jornalismo (UNAMA).

Referências

CARLEY, Kathleen M. Social cybersecurity: an emerging science. **Computational and mathematical organization theory**, v. 26, n. 4, p. 365-381, 2020.

FENG, Shangbin et al. BotRGCN: Twitter bot detection with relational graph convolutional networks. In: **Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining**. 2021. p. 236-239.

NG, Lynnette Hui Xian; CARLEY, Kathleen M. Bot-based emotion behavior differences in images during kashmir black day event. In: **International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation**. Springer, Cham, 2021. p. 184-194.

SAHOO, Somya Ranjan; GUPTA, Brij B. Hybrid approach for detection of malicious profiles in twitter. **Computers & Electrical Engineering**, v. 76, p. 65-81, 2019.

WANG, Yuping et al. Understanding the Use of Fauxtography on Social Media. In: **ICWSM**. 2021. p. 776-786.