

O Papel das Fake News :

Eduardo Kenji Agena

Universidade de São Paulo - USP, São Paulo-SP, Brasil

edukenjister@gmail.com

RESUMO

Devido ao fácil e simples acesso e ao baixo custo e valor agregado, o consumo de notícias online nas mídias sociais (Facebook, Twitter, Instagram, LinkedIn, etc...) aumentou e cresceu significativamente na última década ou últimos dez anos. Apesar de seus benefícios e vantagens, algumas redes sociais (Facebook, Twitter, Instagram, LinkedIn, etc...) permitem que qualquer pessoa ou indivíduo poste notícias online com intenso poder de divulgação e repercussão regional ou mundial, o que amplifica e aumenta um antigo problema: a divulgação ou publicação de Notícias Falsas ou Fake News. Diante desse cenário e situação, vários métodos e processos baseados em aprendizado de máquina para detectar automaticamente Fake News (MLFN) foram propostos e apresentados. Todos eles requerem conjuntos de dados para treinar e avaliar seus modelos de detecção. Embora a MLFN recente tenha sido projetada para considerar dados sobre a propagação de notícias nas mídias sociais (Facebook, Twitter, Instagram, LinkedIn, etc...), a maioria dos poucos conjuntos de dados disponíveis não contém esse tipo de dados. Portanto, comparar os desempenhos entre aqueles MLFN recentes e os outros é restrito a um número muito limitado de conjuntos de dados. Além disso, todos os conjuntos de dados existentes com dados de propagação não contêm notícias em português, o que dificulta a avaliação do MLFN neste idioma. Assim, este trabalho propõe o FakeNewsSetGen, um processo que constrói conjuntos de dados Fake News que contêm dados de propagação de notícias e comparação de suporte em meio ao MLFN de última geração. O método e processo de engenharia de software do FakeNewsSetGen foi orientado para incluir todos os tipos de dados exigidos pelo MLFN existente. Para ilustrar a viabilidade e adequação do FakeNewsSetGen, foi realizado um estudo de caso. Abrangeu a implementação de um protótipo ou projeto-piloto FakeNewsSetGen e a aplicação deste protótipo ou projeto-piloto para criar um conjunto de dados denominado FakeNewsSet, com notícias em português. Cinco MLFN com diferentes tipos de requisitos de dados (dois deles exigindo dados de propagação de notícias) foram aplicados ao FakeNewsSet e comparados, demonstrando o uso potencial do método e processo proposto e do conjunto de dados criado.

Referências

Eduardo Bezerra. 2007. Princípios de Análise e Projeto de Sistemas com UML. Vol. 2. Elsevier, Rio de Janeiro, RJ, Brasil. Google Scholar

C. Buntain e J. Golbeck. 2017. Identificando automaticamente notícias falsas em tópicos populares do Twitter. Em 2017, IEEE International Con on Smart Cloud (SmartCloud). IEEE, New York, NY, USA, 208-215. <https://doi.org/10.1109/SmartCloud.2017.40> Google Scholar

Sonia Castelo et al. 2019. Uma abordagem agnóstica de tópicos para identificar páginas de notícias falsas. Em Companion Proceedings of The 2019 World Wide Web Conference (San Francisco, EUA) (WWW '19). ACM, New York, NY, USA, 975–980. Biblioteca Digital Google Scholar

Perguntas:

- 1)"Qual a principal mensagem do trabalho?" Para responder a esta pergunta, identifique no artigo qual o problema que os autores procuram resolver
- 2)"Qual a principal contribuição do trabalho?" Para responder a esta pergunta, identifique no artigo qual a solução reportada pelos autores
- 3)"Como o trabalho foi avaliado?" Para responder a esta pergunta, identifique no artigo como os autores avaliaram o impacto da solução proposta no problema identificado originariamente
- 4)"Qual a principal limitação do trabalho?" Para a responder esta pergunta, *analise criticamente* o trabalho apresentado no artigo e identifique uma limitação na solução reportada relativamente ao problema tratado

Respostas:

- 1) A disseminação de notícias falsas ou fake news preocupa cada vez mais a sociedade e exige abordagens de mitigação. Embora muitas abordagens tenham sido propostas para a detecção de notícias falsas ou fake news, ainda faltam trabalhos que investiguem profundamente sua estrutura, tempo, veracidade e valor. Nosso estudo foi motivado por duas descobertas discutidas em trabalhos existentes: a primeira é o fato de que notícias falsas geralmente misturam informações reais e falsas para enganar os leitores; a segunda é que a linguagem subjetiva é um recurso comumente explorado por produtores de notícias falsas para fins ideológicos, lucrativos, políticos e partidários;
- 2) Portanto, para entender melhor como as notícias falsas estão estruturadas, os autores realizaram uma análise sobre a forma como a linguagem subjetiva é explorada em diferentes situações dentro dos documentos de notícias falsas ou fake news. Para isso, os autores construíram um conjunto de dados e fatos identificando manualmente as partes falsas de artigos de notícias, e também os autores propõem tags para categorizar os fragmentos em documentos. As tags propostas categorizam os fragmentos de notícias falsas ou fake news de acordo com seu tipo de falsidade no documento ou notícia. Para revelar nuances de subjetividade dentro dos fragmentos, os autores utilizaram o Word Movers Distance e um conjunto de léxicos de subjetividade na língua portuguesa ;
- 3) Os resultados dos autores indicam que a fragmentação das notícias ou documentos permitem a identificação de marcadores de subjetividade que não podem ser identificados ao se considerar a totalidade dos documentos e notícias ;
- 4) Limitações quanto a recuperação e junções das fragmentações de notícias muito antigas e danificadas pelo tempo e má conservação e guarda de documentos para se investigar a veracidade de fatos e dados ocorridos no passado e registrados em jornais, revistas, livros e registros de documentos muito antigos.