

Criação de uma ferramenta para anotação de emoções

Resumo:

Em problemas que utilizam técnicas de aprendizagem de máquina supervisionado é necessário dispor de conjunto de dados para realização de treinamento e teste dos modelos propostos. Estes dados devem estar anotados com as respostas esperadas (gabaritos) e em muitas situações, humanos participam deste processo de anotação de respostas. Neste trabalho foi proposto uma ferramenta Web colaborativa que facilita os usuários e pesquisadores no processo de anotação de emoções em textos, disponibilizando de forma online os textos que devem ser classificados com categorias de emoções pré-definidas. O sistema foi implementado em JavaScript, hospedado na plataforma do Firebase e utilizou o banco de dados Cloud Firestore. Esta ferramenta WEB passou por testes de funcionamento e foi utilizada por diversos usuários, possibilitando anotações de emoções em texto.

Palavras-chave: Ferramenta WEB. Anotação de emoções. JavaScript. Firebase.

Creating a tool for noting emotions

Abstract:

In problems that use supervised machine learning techniques, it is necessary to have a data set for training and testing the proposed models. These data must be annotated with the expected answers (templates) and in many situations, humans participate in this answer annotation process. In this work, a collaborative web tool was proposed that facilitates users and researchers in the process of annotating emotions in texts, providing online texts that must be classified with predefined categories. The system was implemented in JavaScript, hosted on the Firebase platform and used the Cloud Firestore database. This WEB tool has gone through functional tests and has been used by several users, making possible annotations of emotions in text.

Keywords: WEB tool. Emotions annotation. JavaScript. Firebase

Creando una herramienta para notar emociones

Resumen:

En problemas que utilizan técnicas de aprendizaje automático supervisado, es necesario tener un conjunto de datos para entrenar y probar los modelos propuestos. Estos datos deben anotarse con las respuestas esperadas (plantillas) y, en muchas situaciones, los humanos participan en este proceso de anotación de respuestas. En este trabajo se propuso una herramienta web colaborativa que facilita a los usuarios e investigadores en el proceso de anotar emociones en los textos, proporcionando textos online que deben ser clasificados con categorías predefinidas. El sistema se implementó en JavaScript, se alojó en la plataforma Firebase y usó la base de datos de Cloud Firestore. Esta herramienta WEB ha pasado por pruebas funcionales y ha sido utilizada por varios usuarios, possibilitando anotaciones de emociones en texto.

Alexandre Medeiros
Pedra Lavrada, PB, Brasil
IFPB

André Luiz F. Alves
Campina Grande, PB,
Brasil
IFPB

1 Introdução

A análise de sentimentos tem sido uma das áreas mais pesquisadas no campo de Natural Language Processing (LIU, 2015). Descobrir o que as pessoas pensam, segundo Pang e Lee (2008), sempre foi objeto de interesse. No contexto da Web Social, através da popularização de plataformas que fornecem acesso a grande quantidade de dados subjetivos, compreender de forma automática a opinião das pessoas sobre algum tema, serviço ou produto tem sido um fator essencial para a tomada de decisões. Para as organizações, analisar as opiniões das pessoas por meio das mídias sociais significa ampliar as fontes de opinião, tornar mais barata a coleta dos dados e reduzir o tempo de processamento da informação.

As principais abordagens utilizadas para detecção do sentimento estão divididas em quatro técnicas: aprendizagem de máquina (classificação ou regressão); análise léxica (análises linguísticas ou baseadas em dicionários), que utilizam dicionários de palavras com sentimentos já identificados; estatísticas que avaliam a co-ocorrência de termos; e semânticas, que definem a polaridade de palavras em função de sua proximidade semântica com outras de polaridade conhecidas. As diferentes técnicas podem ser combinadas para a obtenção de melhores resultados. As abordagens de aprendizagem de máquina e análise léxica são as mais utilizadas nas aplicações da literatura (LIU, 2015; PANG; LEE, 2008).

Nas técnicas que utilizam aprendizagem de máquina supervisionada, é necessário dispor de conjunto de dados para realização de treinamento e teste dos modelos propostos. Estes dados devem estar anotados com as respostas esperadas (gabaritos) e em muitas situações, humanos participam deste processo de anotação de respostas. No idioma inglês existe uma grande contribuição de dados anotados, possibilitando utilizá-los para aprendizagem de máquina. No idioma português não existem dados suficientes ou de fácil acesso para esse propósito, por isso, foi necessário a criação de uma ferramenta que pode ser utilizada para anotação de quaisquer tipos de dados.

Na área de Processamento de Linguagem Natural (NLP) o reconhecimento de emoções em textos é uma importante tarefa que traz benefícios para diferentes áreas da ciência, como mineração de dados, sistemas de recomendação, educação a distância, interação homem-computador e psicologia (KHATTAK *et al.* 2019). Muitos trabalhos da área de Análise de Sentimentos utilizam as emoções de alegria, raiva, medo, tristeza e

amor. Neste trabalho, além dessas emoções incluímos o rótulo de ironia, para possibilitar maior precisão das anotações dos textos.

2 Metodologia

Nesta seção são descritos os materiais e métodos utilizados. A Figura 1 apresenta as principais etapas desenvolvidas neste trabalho.



Figura 1: Etapas de desenvolvimento do trabalho.

2.1 Tecnologias para o desenvolvimento da ferramenta WEB

O primeiro passo para a criação da aplicação era definir qual linguagem de programação iria usar. A linguagem escolhida foi Javascript, pois é uma linguagem voltada para criação de páginas WEB, além de conter inúmeros *frameworks* e bibliotecas para facilitar todo o processo de desenvolvimento. Foi também utilizado o React.js, que possibilita a criação de interfaces, facilitando a criação de componentes WEB.

Para a hospedagem do site e armazenamento dos dados, foi escolhido o Firebase, que é um banco de dados gratuito, online e servirá como Back-end, facilitando o acesso e resolução de problemas com hospedagem.

2.2 Desenvolvimento da ferramenta WEB

Para desenvolvimento do *Front-end*, o primeiro passo foi realizar a criação de um protótipo da tela inicial, ilustrado na Figura 2, que facilitará todo o processo de criação dos componentes da ferramenta.



Figura 2: Protótipo da ferramenta

Para garantir a identificação dos usuários no sistema, foi necessário implementar um sistema de autenticação. E para evitar que usuários se cadastrassem no sistema,

disponibilizando opções de recuperação senhas, foi decidido utilizar a autenticação dos serviços da google. A autenticação do usuário é feita com o seu e-mail do google por meio do método *firebase.auth()*.

O site foi dividido em 3 telas: a tela de login que faz a autenticação do usuário, a tela inicial do sistema que exibe os dados a serem anotados e os que já foram anotados e uma tela de ajuda, que é apresentado o objetivo da pesquisa, e instruções para o usuário realizar a classificação dos textos de maneira correta.

2.3 Importação dos dados

Os dados importados para o sistema do firebase foram coletados anteriormente por (Lima *et al.* 2020). Esses dados eram tweets que haviam sido coletados e estavam gravados em arquivos CSV e em arquivos JSON. O primeiro passo foi criar um programa em JavaScript para ler esses arquivos, depois foi utilizado o método shuffle para embaralhar e selecionar aleatoriamente 5000 tweets que foram enviados para o Cloud Firestore.

Nesta etapa de importação, salienta-se que qualquer arquivo de texto poderá ser utilizado para o processo de importação dos dados na ferramenta.

2.4 Estruturação dos dados

Os dados foram armazenados no Cloud Firestore que é um banco de dados NoSQL orientado a documentos, que são organizados em coleções. No sistema, foram utilizadas duas coleções: *Users* e *Tweets*.

A coleção *Users* é composta pelas informações dos usuários que realizaram cadastro, cada documento representa um usuário específico, que possui as seguintes propriedades: *createdAt* (armazena a data em que o usuário foi criado), email, nome, sobrenome, *tweetsFeitos* (lista os id's dos tweets que já foram analisados), *tweetsNaoFeitos* (lista os id's que não foram analisados) e *uid* (que é o Id do usuário).

A coleção *Tweets* é composta por informações de cada tweet, seus documentos possuem as seguintes propriedades: *createdAt*, conteúdo (armazena o texto do tweet), *emocao* (é uma propriedade que é subdividida em sete emoções que armazenam a quantidade de vezes em que cada uma foi votada) e *id_str* (é o id do tweet).

3 Resultados e discussão

A Figura 3 apresenta a página inicial da ferramenta que é apresentada para um usuário do sistema. Nela os *tweets* são listados de forma distinta, aleatória e divididos em

blocos, garantindo assim uma maximização na coleta de dados por diversos usuários. Cada bloco representa um tweet, que apresenta o texto a ser analisado ao lado esquerdo, e a direita as 6 emoções definidas previamente (alegria, amor, raiva, medo, tristeza e ironia) que o usuário classificará o texto de acordo com o seu senso crítico. Se não for detectada nenhuma emoção a classificação é rotulada como nula.



Figura 3: Tela de início da ferramenta

A ferramenta foi lançada no dia 09/03/2021 e disponibilizado o endereço para diversos colaboradores participarem de um processo de classificação de emoções. De acordo com os dados obtidos das respostas dos usuários na plataforma foi gerado um gráfico quantitativo para cada emoção, onde 1954 tweets foram analisados pelos usuários. A Figura 4 apresenta o gráfico da porcentagem de vezes que cada emoção foi analisada.

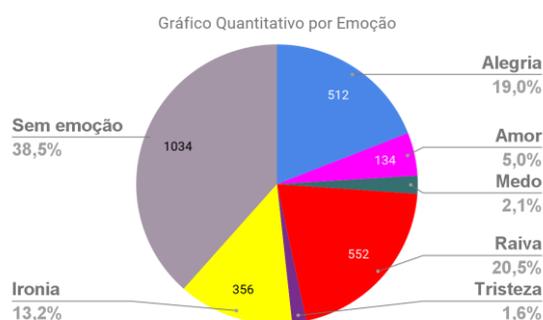


Figura 4: Gráfico quantitativo por emoção

4 Considerações finais

Em técnicas de aprendizagem de máquina supervisionada, é necessário dados rotulados para treinar os modelos de aprendizagem. Neste trabalho foi apresentado uma

ferramenta WEB para anotação de emoções de texto que tem como objetivo facilitar o processo de criação de dados anotados no idioma português para tarefas de Processamento de Linguagem Natural. O conceito da ferramenta é possibilitar a anotação colaborativa de textos, podendo configurar os rótulos desejados e a importação dos textos a serem rotulados. Essa ferramenta já foi utilizada e testada em um processo de classificação de emoções em texto. Deseja-se evoluir no aperfeiçoamento da ferramenta com a possibilidade de adicionar um perfil de usuário para administração da ferramenta, possibilitando bloqueio de usuários e configurações dos textos a serem importados e rótulos das anotações.

Agradecimentos

Os autores agradecem ao CNPq/PIBIC-EM e IFPB por financiar esta pesquisa.

Referências

LIMA, Danielly Rayanne M.; O E SILVA, Raissa Ohana F.; MEDEIROS, Álvaro Getúlio; ALVES, André Luiz. Análise de Sentimentos em Tweets: um Estudo de Caso sobre os cortes de orçamentos nas IFEs. In: ENCONTRO NACIONAL DE COMPUTAÇÃO DOS INSTITUTOS FEDERAIS (ENCOMPIF), 7, 2020, Evento Online. **Anais** [...]. Porto Alegre: Sociedade Brasileira de Computação, 2020. p. 25-28. ISSN 2763-8766.

KHATTAK, Faiza Khan, JEBLEE, Serena, POU-PROM, Chloé, ABDALLA, Mohamed, MEANEY, Christopher, RUDZICZ, Frank. A survey of word embeddings for clinical text. **Journal of Biomedical Informatics**: X, Volume 4, 2019.

PANG, Bo, LEE, Lillian, Opinion Mining and Sentiment Analysis, **Foundations and Trends in Information Retrieval**, Vol 2, Issue 2, 1–135, 2008.

Alexandre Souto Medeiros

Pedra Lavrada, Paraíba, Brasil
IFPB - Campus Picuí

Minicurrículo: Alexandre Souto Medeiros é técnico em informática pelo Instituto Federal da Paraíba, Brasil. Seus interesses de pesquisa são na área de robótica e ciências de dados.

Contribuição de autoria: Escrita – Primeira Redação, Curadoria de dados Metodologia, Software, Validação e Visualização etc.

ORCID ID: <https://orcid.org/0000-0001-6706-965X>

Lattes <http://lattes.cnpq.br/6703057640270975>

E-mail: alexandre.soutomds@gmail.com



André Luiz Firmino Alves

Campina Grande, Paraíba, Brasil
IFPB - Campus Picuí

Minicurrículo: André Luiz F. Alves é Professor do Instituto Federal da Paraíba e doutorando pela Universidade de Campina Grande, Brasil. Ele possui o MSc. e Bel. em Ciência da Computação pela Universidade de Campina Grande, Brasil. Seus interesses de pesquisa são Processamento de Linguagem Natural, com ênfase na análise de sentimentos e emoções.

Contribuição de autoria: Administração do Projeto, Análise Formal, Conceituação, Escrita – Revisão e Edição, Investigação, Metodologia, Obtenção de Financiamento, Recursos, Supervisão, Validação e Visualização etc.

ORCID ID: <https://orcid.org/0000-0001-7883-5129>

Lattes: <http://lattes.cnpq.br/5729800124276465>

E-mail: andre.alves@ifpb.edu.br

