



NOVO MODELO DE RELATÓRIO DOS BOLSISTAS

09/2021

Dados da Bolsa

Tipo de Bolsa: ☒ IC ☐ TCC ☐ PqEP ☐ ME

Nome do/a Orientador/a: Anna Helena Reali Costa

Nome do Projeto: Democracia Aumentada

Período da Bolsa: 02/02/2022 a 01/02/2023

Relatório: ☐ Final ☒ Parcial

Período do Relatório: 25/02/2022 a 25/03/2022

Descrição das Atividades de Pesquisa do Projeto

Descrição das atividades acadêmicas: 2º Semestre Letivo, em que o aluno está cursando as seguintes matérias:

- PEA3306 - Conversão Eletromecânica de Energia
- PEA3301 - Introdução aos Sistemas de Potência
- PEA3311 - Laboratório de Conversão Eletromecânica de Energia
- PTC3307 - Sistemas e Sinais
- PCS3335 - Laboratório Digital A
- PSI3213 - Circuitos Elétricos II
- PEA3100 - Energia, Meio Ambiente e Sustentabilidade
- PMT3100 - Fundamentos de Ciência e Engenharia dos Materiais
- PMT3131 - Química dos Materiais Aplicada à Engenharia Elétrica

Descrição das atividades planejadas para o relatório (repetir do relatório anterior):

Para o próximo mês é planejado encontrar datasets usados como benchmarks dos modelos estado-da-arte

Descrição das atividades de pesquisa realizadas:

Embora para este mês estivesse definido que a atividade seria a busca de novos datasets e benchmarks, com o retorno das atividades presenciais, discutiu-se melhor as opções de foco principal da pesquisa, trabalhando em novas ideias neste mês.



Dentre as ideias, tem-se:

- 1) **Continuar o PLSum** (trabalho de mestrado do C2D): aplicar outras técnicas no sumário capaz de criar um resumo do estilo Wikipédia. O planejamento inicial seria um modelo que replicasse não só a parte do sumário inicial (objetivo original do PLSum), mas também pudesse criar outras partes que comumente aparecem na Wikipedia, no mesmo estilo de linguagem. Referências bibliográficas:
 - a) [Generating PT-BR Wikipedia By Summarizing Multiple Websites - André Seidel](#)
 - b) [The survey: Text generation models in deep learning - Touseef Iqbal](#)
 - c) [Guiding attention in Sequence-to-sequence models for Dialogue Act prediction - Pierre Colombo](#)
 - d) [Topic-Guided Variational Autoencoders for Text Generation - Wenlin Wang](#)
 - e) [Style Example-guided Text Generation Using Generative Adversarial Transformers - Kuo-Hao](#)
 - f) [Towards Data Science - Distill-BERT: Using BERT for Smarter Text Generation](#)
- 2) **Sumarizador Guiado:** Ainda nessa linha de sumarizadores, o foco seria o de criar um modelo próprio de sumarizador guiado. A ideia principal é que este modelo não vise apenas reduzir o volume textual de um documento de entrada, mas sim criar um resumo pautado também no interesse do usuário, dado por especificações extra na entrada do sumarizador, guiando assim a saída do modelo para algo que seja mais específico e representativo. Referências bibliográficas:
 - a) [GSum: A General Framework for Guided Neural Abstractive Summarization](#)
 - b) [Net2Text: Query-Guided Summarization of Network Forwarding Behaviors](#)
 - c) [Fully abstractive approach to guided summarization - Pierre Etienne](#)
 - d) [Keywords-Guided Abstractive Sentence Summarization - Haoran Li](#)
 - e) [SemSUM: Semantic Dependency Guided Neural Abstractive Summarization - Hanqi Jin](#)
 - f) [Guided Summarization for Indonesian News Articles - Danang Massandy](#)
 - g) [Aesthetics-Guided Summarization from Multiple User Generated Videos - Ying Zhang](#)
- 3) **Legendador de Imagens (Text Captioning):** Consiste em juntar técnicas de Processamento de Linguagem Natural e de Visão Computacional para gerar breves descrições das imagens (caption). Esse tipo de modelo é interessante, pois ele usa não só algumas das técnicas já estudadas na IC de geração de textos, mas também novos modelos extratores de features, processamento, reconhecimento e classificação das imagens. Referências bibliográficas:
 - a) [Show, Attend and Tell: Neural Image Caption Generation with Visual Attention](#)
 - b) [Image Captioning - TensorFlow Tutorial](#)
 - c) [Towards Data Science - Image Captioning in Deep Learning](#)
 - d) [An Integrative Review of Image Captioning Research - Chaoyang Wang](#)
 - e) [Every Picture Tells a Story: Generating Sentences from Images - Ali Farhadi](#)
 - f) [A Comprehensive Survey of Deep Learning for Image Captioning - Zakir Hossain](#)
 - g) [Image Captioning using Artificial Intelligence - Yajush Pratap](#)

Ontem (24/03) durante a reunião do grupo foi acordado que os esforços seriam direcionados a expandir o modelo ZeroBERTo no âmbito de:



- Utilizar o modelo em benchmarks conhecidos
- Expandir o modelo atual para comportar múltiplos rótulos (multi-label)
- Aperfeiçoar o modelo para evitar transições de embeddings→palavra (ou vice-versa) desnecessárias
- Aplicar técnicas de *robustness* ao modelo para melhora

Também foi discutido o uso de dados em português, além do uso do Corpus Carolina e do Lexicon, porém estas bases são de um uso mais geral.

Artigos:

Durante o período ocorreu a apresentação do paper “ZeroBERTo: Leveraging Zero-Shot Text Classification by Topic Modeling” no PROPOR2021 que recebeu Menção Honrosa por ser o segundo melhor paper apresentado.

Descrição das próximas atividades:

Para o próximo mês é planejado estudar, implementar e iniciar os testes do ZeroBERTo em outros benchmarks.

Houve alteração significativa no tema ou prazo: () Sim (x) Não

Justifique em caso positivo:

Apreciação Circunstanciada do/a Orientador/a sobre as Atividades da/o Bolsista

Etapa cumprida no relatório: () Ótimo (x) Bom () Regular () Fraco

Programação para a próxima etapa: () Ótimo (x) Bom () Regular () Fraco

Resultados em relação às expectativas iniciais: () Acima (x) Dentro () Abaixo () Muito abaixo

Previsão de conclusão no prazo: (x) Sim () Não

Justifique em caso negativo:

Apreciação da orientadora: O bolsista averiguou e discutiu-se no grupo de estudos diversas alternativas para o prosseguimento da pesquisa, visando soluções mais integradas entre os alunos. Uma alternativa que se mostrou atraente e se decidiu colocar esforços foi o aprimoramento da ferramenta ZeroBERTo, cuja publicação que a descreve foi agraciada com uma Menção Honrosa no evento PROPOR2022. O aluno está motivado e vai se dedicar a isso.



UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA DA UNIVERSIDADE DE SÃO PAULO
DEPARTAMENTO DE ENGENHARIA DE COMPUTAÇÃO E SISTEMAS DIGITAIS
CENTRO DE CIÊNCIA DE DADOS (C²D)



Data: 25/03/2022

Nome Completo da/o Bolsista: Enzo Bustos Da Silva