

Instructions for Authors of SBC Conferences Papers and Abstracts

Luciana P. Nedel¹, Rafael H. Bordini², Flávio Rech Wagner¹, Jomi F. Hübner³

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil

²Department of Computer Science – University of Durham
Durham, U.K.

³Departamento de Sistemas e Computação
Universidade Regional de Blumenau (FURB) – Blumenau, SC – Brazil

{nedel, flavio}@inf.ufrgs.br, R.Bordini@durham.ac.uk, jomi@inf.furb.br

Abstract. This meta-paper describes the style to be used in articles and short papers for SBC conferences. For papers in English, you should add just an abstract while for the papers in Portuguese, we also ask for an abstract in Portuguese (“resumo”). In both cases, abstracts should not have more than 10 lines and must be in the first page of the paper.

Resumo. Este meta-artigo descreve o estilo a ser usado na confecção de artigos e resumos de artigos para publicação nos anais das conferências organizadas pela SBC. É solicitada a escrita de resumo e abstract apenas para os artigos escritos em português. Artigos em inglês deverão apresentar apenas abstract. Nos dois casos, o autor deve tomar cuidado para que o resumo (e o abstract) não ultrapassem 10 linhas cada, sendo que ambos devem estar na primeira página do artigo.

1. Introdução

2. Trabalhos Relacionados

3. Material e Métodos

A metodologia fundamenta-se na premissa de que o viés ideológico se manifesta em padrões discursivos e semântico recorrentes. A abordagem proposta processa o conteúdo textual de notícias utilizando modelos pré-treinados para a geração de *embeddings*, os quais são otimizados via *fine-tuning* com as funções de perda *Contrastive Loss* e *Triplet Loss*. Esse refinamento visa maximizar orientações políticas, servindo como entrada para um classificador de aprendizado de máquina.

Conforme ilustrado na Figura NUMERO, o fluxo de trabalho compreende quatro etapas principais:

- **Aquisição de Dados:** Coleta baseada em conjunto de dados d literatura correlata;
- **Otimização de Representações:** Refinamento de *embeddings* através de aprendizagem métrica (*Contrastive* e *Triplet Loss*);
- **Treinamento:** Calibração do modelo de classificação sobre os vetores otimizados;
- **Avaliação:** análise do desempenho do sistema e dos resultados obtidos.

3.1. Dados Experimentais

3.2. Tarefa de Extração de Representações

3.2.1. Mecanismos de Aproximação e Distanciamento

3.3. Tarefa de Classificação: Modelos e Parametrização

Após o mapeamento dos *embeddings*, onde a proximidade entre os vetores reflete a similaridade ideológica das notícias. A classificação dos artigos foi realizada por meio de três algoritmos: *K-Nearest Neighbors (KNN)*, *K-Means* e *Multilayer Perceptron (MLP)*. O *KNN* e o *K-Means* foram utilizados para explorar a organização dos dados por vizinhança e agrupamento, respectivamente. Para o *KNN*, aplicou-se um *grid search* sistemático para otimização de hiperparâmetros, variando o número de vizinhos (k) entre 5, 10, 15, 20, 25 e 30. Já o *K-Means* foi configurado com o número de *clusters* equivalente às classes presentes no conjunto de dados ABP.

A rede *MLP* foi estruturada com duas camadas densas (512 e 256 neurônios). Adotou-se a função de ativação *ReLU* para garantir um treinamento mais rápido e estável [REFERENCIA], enquanto a camada de saída utilizou a *softmax* para a classificação final. O modelo otimizado com o algoritmo *Adam* [REFERENCIA] e a função de perda *Categorical Cross-Entropy*, escolhas consolidadas na literatura para problemas multiclasse [REFERENCIA GOODFELLOW].

A confiabilidade do experimento foi assegurada pela validacruzada estratificada (5-fold). Esse procedimento garante que a proporção das classes seja mantida em todas as etapas, evitando resultados enviesados e permitindo medir com precisão a capacidade do modelo em classificar novos dados [REFERENCIA]

O desenvolvimento foi implementado em linguagem **Python** utilizando as bibliotecas **Numpy**, **Pandas**, **Scikit-Learn** e **PyTorch**. Os experimentos foram executados em um ambiente de alto desempenho, composto por um processador **Intel Xeon W-2235**, 128 GB de memórias RAM e uma GPU **NVIDIA Quadro TRX 8000** (48 GB), garantindo a eficiência computacional necessária para o treinamento e teste dos modelos propostos. O código-fonte desenvolvido para este estudo, bem como as configurações dos hiperparâmetros e scripts de pré-processamento, estão disponíveis no repositório oficial do projeto: <https://github.com/jailsonpj/detecting-ideological-bias>.

3.4. Avaliação de Desempenho

O desempenho dos modelos de classificação propostos será avaliado quantitativamente através das métricas de Acurácia e *Macro F1-score*. A escolha de tais métricas fundamenta-se na necessidade de uma análise robusta: enquanto a acurácia fornece uma medida geral da taxa de acerto do modelo (Equação 1), o *Macro F1-score* permite uma avaliação criteriosa da capacidade preditiva em todas as classes, mitigando distorções causadas por eventuais desbalanceamentos no conjunto de dados (Equação 2).

As métricas são formalmente definidas como segue:

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Macro } F1 = \frac{1}{N} \sum_{i=1}^N F1_i \quad (2)$$

Onde o $F1$ de cada classe é a média harmônica entre a Precisão e a Revocação, definidas nas Equações 3 e 4:

$$\text{Precisão} = \frac{TP}{TP + FP}, \quad \text{Revocação} = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times (\text{Precisão} \times \text{Revocação})}{\text{Precisão} + \text{Revocação}} \quad (4)$$

Neste contexto, TP , TN , FP e FN representam, respectivamente, os verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos. N indica o número total de categorias ideológicas.

A validação da hipótese de pesquisa — de que o discurso textual reflete o viés ideológico — dar-se-á mediante a obtenção de altos índices em ambas as métricas. Espera-se que valores elevados de *Macro F1-score* confirmem que o modelo possui alta capacidade discriminatória entre as diferentes vertentes ideológicas, garantindo que o desempenho não seja fruto de uma tendência majoritária no conjunto de dados.

4. Resultados e Discussão

5. Considerações Finais

6. References

Bibliographic references must be unambiguous and uniform. We recommend giving the author names references in brackets, e.g. [Knuth 1984], [Boulic and Renault 1991], and [Smith and Jones 1999].

The references must be listed using 12 point font size, with 6 points of space before each reference. The first line of each reference should not be indented, while the subsequent should be indented by 0.5 cm.

Referências

- Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons Ltd.
- Knuth, D. E. (1984). *The TeX Book*. Addison-Wesley, 15th edition.
- Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.