

# Integração de Large Language Models e Graph Neural Networks para Otimizar Classificação de Texto e Extração de Conhecimento

Larissa Menezes  
Fabrício breve  
UNESP  
Rio Claro - SP, Brasil

**Resumo**—Este estudo explora a abordagem híbrida que integra modelos generativos com Redes Neurais Baseadas em Grafos. Embora abordagens tradicionais tenham mostrado sucesso na execução de tarefas de classificação de textos, o foco deste estudo está na integração de modelos generativos e grafos para aprimorar o reconhecimento de entidades e a extração de contexto semântico. Para isso, a pesquisa utiliza o Gemini AI, um modelo de linguagem grande (LLM) desenvolvido pela Google AI, e o REBEL, baseado em BERT, para extrair conhecimento em formato de triplas (entidade, relação, entidade). As triplas extraídas são processadas por uma Rede Neural Convolutiva de Grafos (GCN), que aprimora a compreensão do texto ao propagar o conhecimento das informações dos nós nas camadas da rede. Os resultados preliminares são positivos, destacando o potencial da integração de LLMs e GNNs para a classificação de textos.

*Área: Inteligência Computacional.*

## I. INTRODUÇÃO

A busca por maneiras de representar e estruturar o conhecimento humano precede o surgimento das ferramentas de IA (Inteligência Artificial) modernas. O processamento de linguagem natural (PLN) engloba um conjunto de técnicas e metodologias voltadas para analisar, compreender e extrair significado da linguagem humana de maneira eficaz, permitindo a execução de diversas tarefas, como classificação de texto, atribuição de categorias e rótulos a documentos, refinamento de texto e construção de aplicações assistivas [1]. No campo da PLN, destacam-se os Large Language Models (LLMs), modelos de linguagem treinados com conjuntos de dados de larga escala, treinados para compreender e processar tarefas de linguagem natural [2]. Uma característica vantajosa dos LLMs reside em sua capacidade de se integrar a várias outras abordagens e metodologias no campo da inteligência artificial.

No contexto da IA, a interseção entre grafos de conhecimento e técnicas de aprendizado profundo (*deep learning*) emerge como uma área de pesquisa de grande impacto. Grafos de conhecimento são estruturas utilizadas para representar a relação entre diferentes tipos de entidades (objetos, pessoas,

etc) [3]. Nesse contexto, este estudo explora os conceitos e aplicações de LLMs, grafos de conhecimento e aprendizagem profunda, propondo a criação de um *pipeline* que integra Redes Neurais Baseadas em Grafos (GNNs) com o modelo generativo Gemini AI para automatizar a extração e refinamento de conhecimento, visando aprimorar a classificação de texto. A abordagem consistiu no uso de Graph Convolutional Networks (GCNs) para processar grafos de conhecimento extraídos, com o objetivo de alcançar um desempenho superior a métodos tradicionais de classificação textual.

## II. CONCEITOS E TÉCNICAS

Esse trabalho explora a abordagem híbrida da aplicação de Redes Neurais baseadas em Grafos com modelos generativos para a classificação de textos. Além das técnicas de processamento de linguagem natural, também são utilizados conceitos de identificação e reconhecimento de entidades e contextos semânticos. Os documentos, processados e refinados pelo modelo generativo, são analisados por duas abordagens de extração de conhecimento em triplas (entidade, relação, entidade): uma pelo próprio modelo generativo e outra pelo REBEL (*Relation Extraction By End-to-end Language Generation*) [4], que usa o BERT e um dataset próprio para extrair triplas. Para a etapa de classificação, as representações de conhecimento são processadas por um modelo de Graph Convolutional Networks (GCN), que propaga as informações dos nós através das camadas da rede, enriquecendo a compreensão do conteúdo.

A fase experimental inclui testes com os datasets AG News, uma coleção de artigos de notícias, e Reuters, que contém artigos de diversas categorias. Todos os experimentos foram realizados no Google Colab Pro, com 51 GB de RAM e GPU T4 de 15 GB de RAM.

## III. METODOLOGIA DE DESENVOLVIMENTO

Esta seção descreve as etapas utilizadas na metodologia de desenvolvimento para execução dos experimentos preliminares.

### A. Pré-processamento de dados com Gemini AI

Nessa etapa, foi utilizada a api do Gemini AI para refinar documentos textuais dos datasets escolhidos. Para isso,

utilizou-se um prompt que define os critérios que o modelo deve seguir, como correção de erros gramaticais e ortográficos, mantendo o sentido semântico original do texto.

### B. Extração de conhecimento

A etapa de extração de conhecimento consistiu em converter os documentos refinados pelo Gemini em estruturas textual de grafos de conhecimento em representação de triplas, as quais contém uma coleção de fatos que podem ser denotadas como:  $(h, r, t) \subseteq \varepsilon \times R \times \varepsilon$ , onde  $\varepsilon$  e  $R$  definem as entidades e as relações entre si. A extração se deu por duas abordagens distintas: 1) Utilização do Gemini Api para extrair conhecimento a partir de outro *prompt* pré-definido e, 2) Extração de conhecimento utilizando *REBEL*, um modelo baseado em BERT que identifica entidades e relacionamentos no texto.

### C. Classificação Textual com GCN

Essa etapa consistiu no desenvolvimento da *Graph Convolutional Network (GCN)* em duas abordagens. A primeira envolveu a construção de uma GCN de duas camadas: a camada inicial aplica convolução sobre o grafo, agregando informações dos nós com base nas suas conexões, enquanto a segunda camada refina as representações vetoriais dos nós e gera as pontuações finais para a classificação, com base nas predições. A segunda abordagem adicionou uma terceira camada convolucional ao modelo, visando mapear as representações aprendidas dos nós. Durante a propagação de dados da rede, a primeira abordagem utilizou a função ReLU com *softmax*, e a segunda, a função sigmoide. Em ambas, foi utilizado o *Adam Optimizer* e as funções de perda *Cross-Entropy* e *Binary Cross-Entropy*, conforme definido abaixo:

$$\text{BCE}(y, \hat{y}) = -y \log(\sigma(\hat{y})) - (1 - y) \log(1 - \sigma(\hat{y})) \quad (1)$$

, onde  $y$  são os *labels* verdadeiros e  $\hat{y}$  são as probabilidades preditas

$$\text{CE}(y, \hat{y}) = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (2)$$

, onde  $C$  representa o número de classes,  $y$  os *labels* verdadeiros e  $\hat{y}$  as probabilidades preditas obtidas através da função *softmax*.

## IV. RESULTADOS PRELIMINARES

A comparação dos resultados preliminares é apresentada na Tabela I. Conforme observado, o método Gemini + GCN obteve acurácia de 93% e precisão de 97%, indicando que a maioria das predições foram corretas e, quando positivas, quase sempre precisas. Além disso, os testes de validação com o dataset AG News mostraram um valor de *loss* mais baixo, conforme Figura 1. O segundo método (Gemini AI + REBEL + GCN) obteve acurácia de 92% e precisão de 83%, sendo ligeiramente inferior ao primeiro, com valores elevados de *loss*, sugerindo a necessidade de ajuste de hiperparâmetros e aprimoramento do treinamento.

Tabela I  
PERFORMANCE DO MODELO DE CLASSIFICAÇÃO TEXTUAL

Dataset	Método	Épocas	Loss	Acurácia	Precisão
AG News	Gemini AI + GCN	1000	0.023	0.93	0.97
Reuters	REBEL + GCN	1000	0.68	0.92	0.83

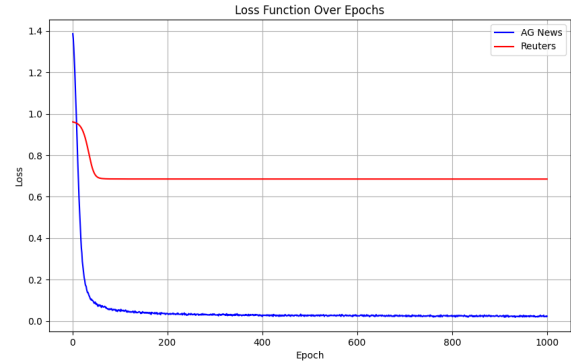


Figura 1. Valores de *loss* através das épocas

## V. CONSIDERAÇÕES FINAIS

Este estudo apresentou uma abordagem híbrida que combina Large Language Models (LLMs) com Redes Neurais Baseadas em Grafos (GNNs) para aprimorar a classificação de textos e a extração de conhecimento. A integração dos modelos Gemini AI e REBEL, com o uso de Graph Convolutional Networks (GCNs), demonstrou um potencial significativo na melhoria do desempenho de tarefas de classificação textual. Os próximos passos envolvem a realização de testes com outros datasets, para avaliar a generalização do modelo. Além disso, será feito o aprimoramento das redes neurais, incluindo ajustes de hiperparâmetros e o uso de técnicas de regularização. Também será conduzida uma análise comparativa com outros modelos disponíveis.

## REFERÊNCIAS

- [1] A. Mansi, "An overview of natural language processing," vol. 7, no. 5, pp. 2811–2813.
- [2] H. Naveed, A. U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman, N. Barnes, and A. S. Mian, "A comprehensive overview of large language models," *ArXiv*, vol. abs/2307.06435, 2023. [Online]. Available: <https://api.semanticscholar.org/CorpusID:259847443>
- [3] Z. Ye, Y. Jaya Kumar, G. Sing, F. Song, and J. Wang, "A comprehensive survey of graph neural networks for knowledge graphs," *IEEE Access*, vol. 10, pp. 1–1, 01 2022.
- [4] P.-L. Huguet Cabot and R. Navigli, "REBEL: Relation extraction by end-to-end language generation," in *Findings of the Association for Computational Linguistics: EMNLP 2021*, M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, Eds. Punta Cana, Dominican Republic: Association for Computational Linguistics, Nov. 2021, pp. 2370–2381. [Online]. Available: <https://aclanthology.org/2021.findings-emnlp.204>