

AGREGAÇÃO DE PREDIÇÕES DE GRANDES MODELOS DE LINGUAGEM VIA MÉTODOS DE DECISÃO EM GRUPO PARA A CODIFICAÇÃO AUTOMÁTICA DE DIAGNÓSTICOS CLÍNICOS

Ricardo da Silva Santos

IMECC/UNICAMP

Rua Sérgio Buarque de Holanda, 651, Cidade Universitária Zeferino Vaz. Campinas, SP

`ricardo.santos@ime.unicamp.br`

Murilo Gleyson Gazzola

FCI/MACKENZIE

Av. Mackenzie, 905, Tamboré. Barueri, SP

`murilo.gazzola@mackenzie.br`

Renato Teixeira Souza

FCM/UNICAMP

Rua Vital Brasil, 80. Cidade Universitária Zeferino Vaz. Campinas, SP

`rtsouza@g.unicamp.br`

Rodolfo de Carvalho Pacagnella

FCM/UNICAMP

Rua Vital Brasil, 80. Cidade Universitária Zeferino Vaz. Campinas, SP

`rodolfofop@unicamp.br`

Cristiano Torezzan

FCA/UNICAMP

Rua Pedro Zaccaria, 1300, Limeira, SP

`torezzan@unicamp.br`

RESUMO

Grandes modelos de linguagem (LLMs) vêm sendo utilizados na codificação automática de diagnósticos clínicos, traduzindo descrições textuais em códigos da Classificação Internacional de Doenças (CID-10). Contudo, diferentes arquiteturas e configurações de LLMs frequentemente produzem predições divergentes para um mesmo caso, gerando incertezas sobre a escolha mais adequada. Este trabalho propõe uma abordagem baseada em métodos de decisão em grupo para consolidar as predições geradas por múltiplos LLMs. Cada modelo é tratado como um decisor que fornece uma ordenação dos códigos mais prováveis, e técnicas de agregação, como o método de Borda e variantes de votação condorcetiana, são aplicadas para identificar os códigos de maior consenso. A abordagem é avaliada com dados reais de notas clínicas de alta hospitalar em um hospital especializado em saúde materno-neonatal. Os resultados indicam que a agregação das predições melhora significativamente a precisão da codificação, reduz a variabilidade entre modelos e supera o desempenho individual dos LLMs. A proposta evidencia o potencial dos métodos de decisão em grupo como metamodelos robustos para tarefas complexas de classificação, promovendo a integração entre inteligência artificial e apoio à decisão em sistemas de saúde.

PALAVRAS CHAVE. Codificação automática de diagnósticos, Grandes modelos de linguagem, Métodos de decisão em grupo

Tópicos: ADM – Métodos Multicritério para Ajuda de Tomada de Decisão; SA – PO na Área de Saúde; IA – PO e IA.

ABSTRACT

Large language models (LLMs) have been used for the automatic coding of clinical diagnoses, translating textual descriptions into codes from the International Classification of Diseases (ICD-10). However, different LLM architectures and configurations often produce divergent predictions for the same case, generating uncertainty regarding the most appropriate code. This work proposes an approach based on group decision-making methods to consolidate predictions generated by multiple LLMs. Each model is treated as a decision-maker that provides a ranking of the most likely codes, and aggregation techniques such as the Borda count and variants of Condorcet voting are applied to identify the codes with the highest consensus. The approach is evaluated using real hospital discharge summaries from a maternity and neonatal health care facility. The results indicate that aggregating predictions significantly improves coding accuracy, reduces inter-model variability, and outperforms individual LLMs. The findings highlight the potential of group decision-making methods as robust metamodels for complex classification tasks, fostering integration between artificial intelligence and decision support in healthcare systems.

KEYWORDS. Automatic Diagnosis Coding; Large Language Models; Group Decision-Making Methods

Paper Topics: ADM – Multicriteria Decision-Making Methods; SA – OR in Healthcare; IA – OR and AI.