

Rotulação de Exercícios da Olimpíada Brasileira de Informática (OBI) com Ênfase em Aplicações Educacionais

Felipe Jun Takahashi¹, Mateus de Oliveira Sposo¹

¹Universidade Federal de Mato Grosso do Sul (UFMS)
Campo Grande – MS – Brazil

{felipe.takahashi, mateus_sposo}@ufms.br

Abstract. *In recent years, academic research has shown a growing interest in applying Artificial Intelligence to education. This paper presents the manual classification of programming exercises from the Brazilian Olympiad in Informatics (OBI) for 2023 and 2024. For a total of 132 instances, we analyzed topic frequency and the mapping of topics by level and stage. Finally, we discuss practical applications in the Faculty of Computing's (FACOM) competitive programming education program, FACOMpetindo.*

Resumo. *Nos últimos anos, pesquisas acadêmicas demonstraram um interesse crescente em Inteligência Artificial aplicada à educação. Este artigo descreve a classificação manual de exercícios de programação da Olimpíada Brasileira de Informática (OBI) de 2023 e 2024. Em um total de 132 instâncias, analisamos a frequência dos tópicos e o mapeamento dos tópicos por nível e fase. Por fim, discutimos aplicações práticas no programa de educação em programação competitiva da Faculdade de Computação (FACOM), FACOMpetindo.*

palavras-chave – Educação, Olimpíada, Ciência da Computação, OBI, Inteligência Artificial, Classificação, Rotulação

1. Introdução

Em razão do rápido crescimento da Inteligência Artificial e Big Data, o cenário acadêmico explora diferentes modos de auxiliar docentes e estudantes por meio do estudo de áreas como *Educational Data Mining* e *Learning Analytics*¹. Diversas aplicações educacionais (sistemas de recomendação, preditores de desempenho, identificadores de fraquezas) fundamentam-se nesses tópicos. Nesse sentido, a categorização de dados é um processo imprescindível para o desenvolvimento de sistemas pedagógicos sofisticados [Bienkowski et al. 2014].

No contexto da educação, destacam-se iniciativas como a Olimpíada Brasileira de Informática (OBI) [Anido and Menderico 2007], que explora conceitos matemáticos e computacionais fundamentais ao ensino de computação. Com ênfase na modalidade programação, a estrutura dos cadernos de prova possibilita o agrupamento de questões em categorias referentes ao tema. Mais que isso, a existência de níveis (por idade) e fases (por desempenho) permite a separação dos exercícios de acordo com o público-alvo.

Nesse sentido, este relatório apresenta os resultados de um estudo inicial, a rotulação de exercícios da OBI (modalidade programação), com o objetivo de apoiar o

¹Mais de 50 mil publicações desde 2020, pelo Google Scholar (busca em maio de 2025)

ensino de computação e investigar métodos de desenvolvimento de sistemas educacionais inteligentes.

2. Metodologia

2.1. Conjunto de Dados e Critérios

A partir dos cadernos de prova, soluções e gabaritos disponibilizados no site oficial da olimpíada ², foram rotulados um total de 132 exercícios, ou 72 excluindo múltiplas instâncias. Cada exercício é composto por cinco atributos: *Nome*, *Nível*, *Fase*, *Ano*, e *Rótulos*, com o intuito de extrair e definir padrões temáticos existentes na prova. Como critérios para a rotulação, foi estabelecido que todo problema deve conter ao menos um rótulo e, em caso de múltiplos tópicos, todos são contabilizados. Para evitar ambiguidades, todo rótulo é definido por um critério objetivo e não ambíguo. (1).

Tabela 1. Descrição dos critérios de classificação por rótulo

Rótulo	Descrição
Ad Hoc	A solução do exercício é baseada, exclusivamente, em regras específicas ao enunciado.
Estrutura de Dados	A resolução do exercício destaca estruturas de dados como vetores, listas, filas, pilhas, conjuntos, árvores ou grafos.
Força Bruta	O exercício é resolvido por técnicas de força bruta, como enumeração de possibilidades.
Geometria Computacional	A abordagem aplica conceitos e algoritmos de geometria computacional.
Guloso	A solução destaca escolhas locais ótimas com comportamento guloso.
Matemática	A solução requer a aplicação direta de conceitos matemáticos, como aritmética, álgebra, geometria ou teoria dos números.
Ordenação	A solução do exercício depende de princípios e algoritmos de ordenação.
Programação Dinâmica	O exercício demanda técnicas de programação dinâmica.
Strings	A solução do exercício enfatiza a manipulação e processamento de strings.
Teoria dos Grafos	A solução envolve conceitos e algoritmos próprios de grafos.

2.2. Classificação Manual

Em primeira instância, optou-se por um processo manual para determinar a qualidade dos atributos definidos. Nesse âmbito, todas as questões do ano de 2023 e 2024 foram classificadas manualmente, abrangendo todos os níveis (Júnior, 1, 2, Sênior), fases (1, 2, 3, Única) e turnos (A, B) na modalidade programação. Durante o processo, nenhum conjunto de problemas apresentou ambiguidade ou ausência de rótulos para seus elementos, destacando a clareza dos critérios adotados. Nessa perspectiva, a metodologia apresenta potencial de replicação ou extensão em contextos similares.

²<https://olimpiada.ic.unicamp.br/>

2.3. Processo de Classificação

Durante o processo de classificação manual, os dados foram rotulados em uma planilha no Google Sheets e processados por scripts simples em Python. A atribuição de rótulos a cada exercício baseou-se na descrição fornecida pelo caderno de provas e, em seguida, validada com auxílio das soluções públicas disponibilizadas no site da OBI, além de uma revisão final pelos autores. Além disso, para exercícios idênticos em níveis distintos, todas as instâncias foram individualmente classificadas e consideradas na análise.

2.4. Classificação Automática

Por fim, em razão da ausência de conjuntos previamente rotulados para fins de validação, decidiu-se por postergar a classificação automatizada. Ademais, a automatização do processo em estágio inicial poderia comprometer a avaliação da qualidade dos atributos, em especial, dos rótulos escolhidos para a rotulação. Diante do potencial evidenciado pela categorização manual, planeja-se, em etapas futuras, a implementação de um agente classificador, potencialmente com base em técnicas de *zero-shot learning*, utilizando os dados formulados neste processo inicial.

3. Resultados

Após a categorização, destacou-se a semelhança estrutural da OBI com outras provas de programação competitiva (e.g., ICPC³, Codeforces⁴, IOI⁵). Estudos realizados sobre modelos *zero-shot* indicam que exercícios de competições são um conjunto de dados efetivos para avaliar *Large Language Models* (LLMs) [Huang et al. 2023b]. O baixo desempenho de LLMs, especialmente em raciocínio lógico, em idiomas diferentes do inglês [Lai et al. 2023] motiva a criação de *benchmarks* que contemplam múltiplas línguas para avaliar os modelos.

No contexto de classificação, a separação entre níveis e fases na prova torna o conjunto bem distribuído em dificuldades e complexidade linguística. Contudo, a atual discrepância na frequência de exercícios por tópico impede a formulação de um *benchmark* consistente para modelos classificadores nesse estágio.

3.1. Frequência de Tópicos

Ao analisar a frequência de tópicos no conjunto, observa-se que *Estrutura de Dados* e *Matemática* são os rótulos mais recorrentes, 49 (30,8%) e 31 (19,5%) ocorrências respectivamente. Os rótulos de menor frequência são *Geometria Computacional*, 4 (2,5%) ocorrências, e *Força Bruta*, 2 (1,3%) ocorrências.

Apesar de uma variação em complexidade, os dados mostram a concentração em tópicos fundamentais da computação. Em contrapartida, técnicas especializadas ocorrem em menor frequência. Além disso, é possível notar uma ocorrência intermediária de *Teoria dos Grafos* (10,7%) e *Ad Hoc* (9,4%), que se assemelham às habilidades exploradas em *Estrutura de Dados* e *Matemática*.

³<https://icpc.global/>

⁴<https://codeforces.com/>

⁵<https://ioinformatics.org/>

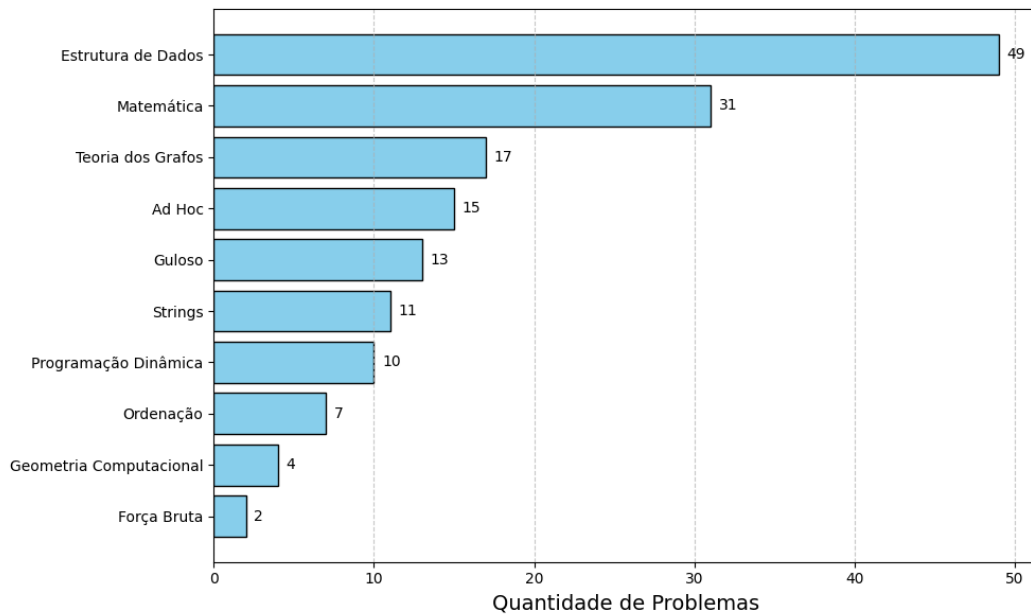


Figura 1. Quantidade de problemas por tópico

3.2. Distribuição em Níveis

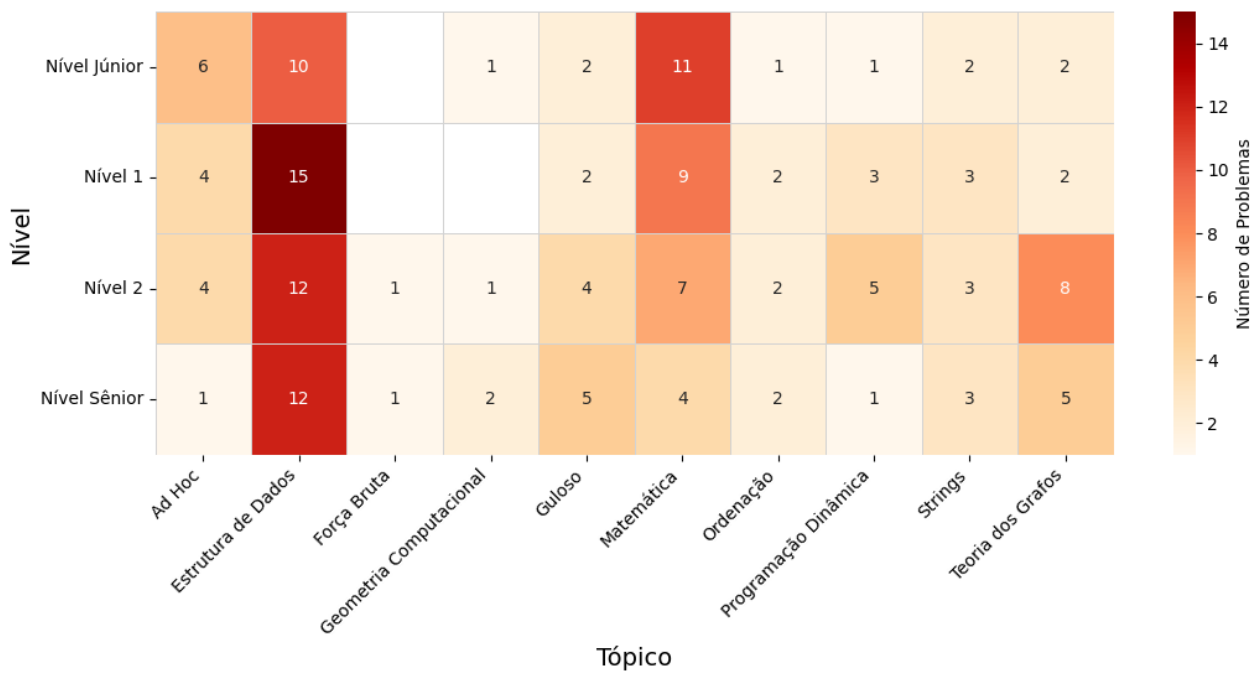


Figura 2. Cobertura de tópicos por nível

A cobertura por níveis destaca os tópicos mais frequentes em um determinado nível, independente da fase. Neste caso, observa-se que *Estrutura de Dados* é predominante em todos os níveis. Entretanto, *Ad Hoc*, *Guloso* e *Teoria dos Grafos* são relevantes em grupos específicos. No contexto de múltiplas instâncias de um exercício, a distribuição enfatiza a similaridade entre os níveis Júnior e 1, de forma análoga, os níveis 2 e Sênior apresentam a mesma similaridade, compartilhando mais questões entre si do que com outros níveis.

3.3. Distribuição em Fases

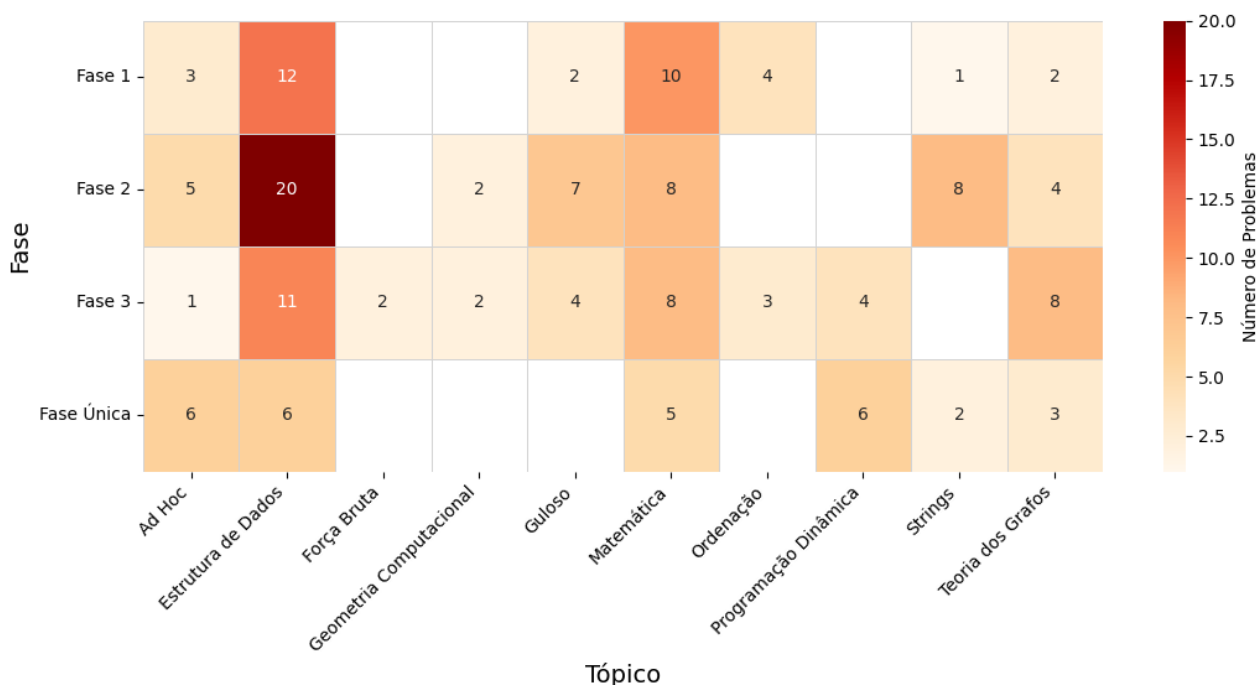


Figura 3. Cobertura de tópicos por fase

Similar à divisão por níveis, a progressão em fases permite a distinção dos exercícios por dificuldade. Em especial, o mapa ressalta o conhecimento esperado em cada fase. Na etapa inicial, a variedade de rótulos é menor do que em outras etapas, destacando o foco em conceitos fundamentais. Conforme a progressão de fases, tópicos antigos são aprofundados, e novos conteúdos são adicionados. Problemas de *Estrutura de Dados* na Fase 3 são conceitualmente mais ricos que na Fase 2 que, por sua vez, exigem mais do que a Fase 1 (e.g., estruturas específicas, manipulações mais complexas, casos de teste mais exigentes).

3.4. 2023 & 2024

Nos dois anos classificados, notou-se que existem padrões persistentes em fases e níveis específicos. Os níveis Júnior e 1 compartilham quase todas as questões em fases finais,

com uma diferença de, no máximo, duas questões na terceira fase. O mesmo comportamento ocorre para os níveis 2 e Sênior. Em destaque, a Fase 1 apresenta a menor diversidade temática e o menor número de questões exclusivas por nível. Enquanto isso, a última fase apresenta a maior quantidade de instâncias repetidas, e a maior variedade de conteúdos.

4. Aplicações na Educação e Interesses

4.1. FACOMpetindo

O programa de ensino em programação competitiva, FACOMpetindo, realiza aulas semanais para todos os alunos da Faculdade de Computação (FACOM) com interesse em programação competitiva. Como resultado direto do processo de classificação manual, os dados coletados serão fornecidos aos instrutores para o auxílio na criação de material de estudo, baseado na OBI, para a preparação de calouros.

4.2. Classificador automático

Nesta primeira etapa, apenas os exercícios de 2023 e 2024 foram classificados. Entretanto, a prova é realizada anualmente e o site da olimpíada disponibiliza os cadernos de prova desde 2001. Com a construção de um conjunto rotulado manualmente, planeja-se continuar o estudo através da automação do processo por meio de técnicas de *zero-shot* e *few-shot* learning. Particularmente, resultados de pesquisas que implementaram modelos similares para a classificação de problemas de programação, em inglês, exploram o potencial da automação [Pereira et al. 2021, Kim et al. 2024].

4.3. Sistema de recomendação

Em meio as diversas aplicações de inteligência artificial na educação (e.g., criação automática de exercícios, detecção de plágio, previsão de desempenho) [Yang et al. 2021, Zhong et al. 2019, Rivera et al. 2018], destaca-se a criação de sistemas de recomendação para guiar o estudo de um aluno. Através de análises no engajamento de estudantes e a implementação de um sistema de recomendação inteligente em uma plataforma de vídeos, um estudo conduzido na China observou que as recomendações são capazes de ampliar a motivação e o desempenho dos acadêmicos, especialmente de indivíduos moderadamente engajados no aprendizado [Huang et al. 2023a].

Nesse sentido, propõe-se a preparação de dados para a criação de um sistema de recomendação baseado em exercícios da OBI. Em especial, indica-se a colaboração com o projeto educacional, FACOMpetindo, respeitando os princípios éticos na coleta e uso de informações [Floridi and Taddeo 2016], para a representação de perfis dos estudantes, além da participação ativa dos monitores para validar as sugestões geradas [Isinkaye et al. 2015].

5. Conclusão

As características estruturais da prova descobertas nesse estudo amplificam a capacidade de criação de material didático para estudantes. O compartilhamento de exercícios entre níveis faz possível a separação de aulas em dois grupos (Júnior e 1; 2 e Sênior), ao invés de quatro níveis (Júnior; 1; 2; Sênior). Além disso, a criação de um conjunto de dados rotulado manualmente viabiliza a implementação de um modelo de classificação

automática aplicável a anos anteriores e futuros. Por fim, idealiza-se a criação de um sistema de recomendação de exercícios em uma aplicação educacional, para auxiliar alunos e docentes ao longo do processo de aprendizagem.

Referências

- Anido, R. d. O. and Menderico, R. M. (2007). Brazilian olympiad in informatics. *Olympiads in Informatics*, 1:5–14.
- Bienkowski, M., Feng, M., and Means, B. (2014). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.
- Floridi, L. and Taddeo, M. (2016). What is data ethics? *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences*, 374:20160360.
- Huang, A. Y., Lu, O. H., and Yang, S. J. (2023a). Effects of artificial intelligence-enabled personalized recommendations on learners' learning engagement, motivation, and outcomes in a flipped classroom. *Computers Education*, 194:104684.
- Huang, Y., Lin, Z., Liu, X., Gong, Y., Lu, S., Lei, F., Liang, Y., Shen, Y., Lin, C., Duan, N., and Chen, W. (2023b). Competition-level problems are effective llm evaluators.
- Isinkaye, F., Folajimi, Y., and Ojokoh, B. (2015). Recommendation systems: Principles, methods and evaluation. *Egyptian Informatics Journal*, 16(3):261–273.
- Kim, J., Cho, E., and Na, D. (2024). Problem-solving guide: Predicting the algorithm tags and difficulty for competitive programming problems.
- Lai, V. D., Ngo, N. T., Veyseh, A. P. B., Man, H., Deroncourt, F., Bui, T., and Nguyen, T. H. (2023). Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning.
- Pereira, F. D., Pires, F., Fonseca, S. C., Oliveira, E. H. T., Carvalho, L. S. G., Oliveira, D. B. F., and Cristea, A. I. (2021). Towards a human-ai hybrid system for categorising programming problems. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education, SIGCSE '21*, page 94–100, New York, NY, USA. Association for Computing Machinery.
- Rivera, A. C., Tapia-Leon, M., and Lujan-Mora, S. (2018). Recommendation systems in education: A systematic mapping study. In Rocha, Á. and Guarda, T., editors, *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)*, pages 937–947, Cham. Springer International Publishing.
- Yang, S. J., Ogata, H., Matsui, T., and Chen, N.-S. (2021). Human-centered artificial intelligence in education: Seeing the invisible through the visible. *Computers and Education: Artificial Intelligence*, 2:100008.
- Zhong, J., Xie, H., and Wang, F. L. (2019). The research trends in recommender systems for e-learning. *Asian Association of Open Universities Journal*, 14(1):12–27.