

Descrição do seminário da disciplina Aprendizado Descritivo 2025/1

Renato Vimieiro

2025-05-17

Introdução

O seminário consistirá em uma apresentação coletiva (da turma) de artigos mais recentes sobre tópicos diretamente relacionados ao conteúdo visto em sala. A ideia é que a turma como um todo estude os artigos mais recentes da área e discuta esses trabalhos em sala. Em cada sessão, discutiremos dois artigos recentes relacionados aos tópicos abordados nas aulas teóricas.

Para isso, adotaremos um formato adaptado da proposta apresentada pelos Profs. Alec Jacobson e Colin Raffel, ambos da Universidade de Toronto (Canadá) – veja a proposta original em <https://colinraffel.com/blog/role-playing-seminar.html>. A proposta consiste em fazer uma encenação de papéis (role play) científicos para a apresentação do seminário. Nessa proposta, cada grupo (a proposta original previa que cada aluno assumisse um dos papéis, o que é inviável para o tamanho da nossa turma) cumprirá um papel na apresentação. A lista de papéis é apresentada a seguir:

1. Historiador: deve contextualizar o problema; identificar que tipo de padrão visto em aula melhor se adequa ao abordado no artigo; em suma, deve relacionar o contexto, motivação, justificativa e pergunta de pesquisa do artigo com os temas vistos em aula, deixando clara as relações (similaridades e diferenças) entre os dois. Se possível, citar métodos clássicos similares ao que é apresentado no artigo.
2. Formalizador: deve apresentar formalmente os padrões minerados no artigo e demais conceitos chave. Esse papel é responsável por mostrar os conceitos fundamentais do artigo que possibilitem compreender plenamente o método/aplicação discutido no trabalho. No caso de apresentação de novos algoritmos, esse também deverá ser apresentado de maneira mais didática à turma. Se possível, tente simplificar a linguagem utilizada, tornando o artigo mais compreensível ao público amplo (o formalismo deve ser compatível com o de alunos da graduação a partir do 6º período, ou seja, que já tenham cursado disciplinas de algoritmos, matemática discreta e lógica).

3. Metodologista: deve apresentar a metodologia usada no artigo para comparar o método, validar a hipótese do artigo. Esse papel é responsável por analisar os experimentos conduzidos no artigo, apresentando-os para a turma. Deve-se observar qual foi a estratégia usada para realizar os experimentos, quais eram as hipóteses/perguntas de pesquisa que estavam sendo respondidas com cada experimento, o uso ou não de ferramentas estatísticas (por exemplo, testes de hipótese) para validar os achados. Finalmente, o grupo também deve fazer uma análise crítica dos resultados e conclusões do trabalho apresentados pelos autores.
4. Assessor social: deve avaliar o impacto social (aplicações, produtos etc.) do trabalho ou de temas relacionados. Se possível, apresentar cenários para além dos mostrados no artigo em que a ferramenta seria útil. O grupo também deve tentar prever problemas com aplicações envolvendo o método em relação a aspectos de privacidade, equidade, discriminação, igualdade, ou outros possíveis danos à sociedade que o método (ou seu uso) pode trazer.
5. Hacker: é responsável por verificar a existência de repositórios com o código fonte do artigo, coletar bases de dados para exemplificar os padrões encontrados pelo método. O grupo deve criar o roteiro para baixar, instalar e rodar o método localmente. Deve descrever o formato de entrada dos dados, caso seja algum formato especial (não seja csv, json, arff).
6. Relator: relata as apresentações do dia e prepara um documento agregando o que foi discutido. Note que, enquanto os papéis anteriores têm como objetivo construir uma apresentação, esse papel é responsável por documentar textualmente o conteúdo da aula. O relator é responsável por formatar e divulgar o relatório na página coletiva da disciplina (que será tornada pública para a comunidade). A intenção é fazermos uma espécie de divulgação científica do tema abordado no artigo. O alvo será o público especializado de computação/ciência de dados, mas devemos estruturar a nossa publicação de maneira que seja compreensível para pessoas com diferentes níveis de formação nessas áreas. A equipe deverá compilar possíveis imagens, tabelas e informações não textuais dos outros grupos para incrementar o relatório do grupo anterior. O grupo também deve se atentar para a diagramação do conteúdo, tentando torná-lo mais agradável ao público alvo. Esses são alguns exemplos de resultado para essa tarefa:
 - <https://medium.com/@abnersuniga7/encontre-padrões-nos-seus-dados-com-apriori-e-fp-growth-4a581ec1b22>
 - <https://towardsdatascience.com/frequent-pattern-mining-association-and-correlations-8fa9f80c22ef>
 - <https://medium.com/@ciortanmadalina/an-introduction-to-frequent-pattern-mining-research-564f239548e#:~:text=Frequent%20pattern%20mining%20is%20a,making%20C%20clustering%20website%20navigation>.

Note que a linguagem utilizada é bastante informal, mas ainda se percebe diversos elementos científicos, por exemplo, citações para trabalhos usados como referência na escrita do texto.

A turma será dividida em seis grupos de maneira arbitrária pelo professor. A lista com a associação dos nomes e grupos (papéis) estará no Moodle.

O que deverá ser entregue?

As equipes de 1 a 5 (inclusive) deverão fazer uma apresentação com o conteúdo associado ao seu papel. Os slides dessa apresentação serão compartilhados entre todas as equipes, sendo cada uma responsável por incluir os slides referentes à sua parte. O link para o slide compartilhado será divulgado posteriormente. Dessa forma, toda a apresentação do artigo estará centralizada em um único arquivo, o qual será projetado pelo professor durante as aulas de apresentação dos seminários. A razão é diminuir o tempo necessário entre as trocas de grupos a apresentar e possíveis problemas técnicos. Além disso, o material centralizado facilitará a construção do relatório e demais publicações dos grupos restantes.

Como teremos apresentações de 5 grupos para cada artigo, haverá limite de 4 slides para cada grupo, limitando o tempo de apresentação em 15min. Lembrem-se que as apresentações devem ser focadas especificamente no tema associado ao papel da equipe. Outros aspectos que o grupo julgue ser importante para compreender o conteúdo do seu tema, mas que esteja relacionado a outro grupo, poderá ser sugerido para inclusão ao grupo responsável. Haverá um canal no Moodle para postagens com sugestões de assuntos aos demais grupos para cada artigo/seminário. Mais uma vez, a intenção é trabalharmos em conjunto para compreendermos melhor o conteúdo dos artigos. Então é extremamente importante o engajamento nas tarefas com sugestões e comentários para os colegas.

As apresentações do artigo da sessão devem ser entregues até às 12h00 (meio-dia) do dia da apresentação.

O relatório a ser produzido pela equipe responsável será escrito em documento compartilhado (Google Docs) com a turma. No dia da apresentação, o grupo deverá escrever suas notas em materiais próprios, e, no prazo de 10 dias corridos após a apresentação do artigo, deverá ter concluído a escrita do relatório final no arquivo compartilhado e a publicação da página como descrito anteriormente. Para a divulgação, usaremos o sistema de publicação Quarto (<https://quarto.org>). Esse modelo usa uma linguagem de marcação para formatar as páginas. Essas serão armazenadas em um repositório público do Github. Finalmente, a página será automaticamente gerada quando houver o push para o repositório. Os alunos serão incluídos como colaboradores do repositório, portanto, terão autonomia para realizar commits e pushes necessários. A avaliação do trabalho será com base no último commit feito no dia da entrega.

Artigos

Os artigos abordados em cada sessão estão listados abaixo. As sessões foram divididas conforme o tema do seminário. Teremos três sessões com dois artigos em cada.

- Primeira sessão (mineração de padrões)
 1. Fischer and Vreeken (2021)
 2. Ouarem, Nouioua, and Fournier-Viger (2024)
- Segunda sessão (padrões supervisionados)
 1. Walter, Fischer, and Vreeken (2024)
 2. Heinrich et al. (2025)
- Terceira sessão (aplicações)
 1. Leeuw et al. (2022)
 2. Schouten et al. (2024)

Referências

- Fischer, Jonas, and Jilles Vreeken. 2021. “Differentiable Pattern Set Mining.” In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, 383–92. KDD ’21. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3447548.3467348>.
- Heinrich, Carl Vico, Tommie Lombarts, Jules Mallens, Luc Tortike, David Wolf, and Wouter Duivesteijn. 2025. “Local Subgroup Discovery on Attributed Network Graphs.” In *International Symposium on Intelligent Data Analysis*, 195–208. Springer.
- Leeuw, Arie-Willem de, Rick van Baar, Arno Knobbe, and Stephan van der Zwaard. 2022. “Modeling Match Performance in Elite Volleyball Players: Importance of Jump Load and Strength Training Characteristics.” *Sensors* 22 (20). <https://doi.org/10.3390/s22207996>.
- Ouarem, Oualid, Farid Nouioua, and Philippe Fournier-Viger. 2024. “Discovering Frequent Parallel Episodes in Complex Event Sequences by Counting Distinct Occurrences.” *Applied Intelligence* 54 (1): 701–21. <https://doi.org/10.1007/s10489-023-05187-y>.
- Schouten, Rianne M., Wouter Duivesteijn, Pekka Räsänen, Jacob M. Paul, and Mykola Pechenizkiy. 2024. “Exceptional Subitizing Patterns: Exploring Mathematical Abilities of Finnish Primary School Children with Piecewise Linear Regression.” In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track*, edited by Albert Bifet, Tomas Krilavičius, Ioanna Miliou, and Slawomir Nowaczyk, 66–82. Cham: Springer Nature Switzerland.
- Walter, Nils Philipp, Jonas Fischer, and Jilles Vreeken. 2024. “Finding Interpretable Class-Specific Patterns Through Efficient Neural Search.” In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI’24/IAAI’24/EAAI’24. AAAI Press. <https://doi.org/10.1609/aaai.v38i8.28756>.