

abril/2023 • n. 49

COMPUTAÇÃO

REVISTA DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO — BRASIL

Mineração de
Processos:
do que se trata?
E o Brasil, está no jogo?

EDITORIAL

Este ano, a SBC comemora 45 anos de existência, um marco histórico que será celebrado em um evento especial que revisitará as realizações marcantes de nossa sociedade.

A importância estratégica da SBC como sociedade científica na área de computação no Brasil é evidente, já que foi a principal responsável pelo estabelecimento da carreira e da formação acadêmica nesse campo, por meio de referenciais curriculares e outras iniciativas. Ao longo desses anos, a SBC tem desempenhado um papel fundamental na atualização de conhecimentos para pesquisadores, docentes, profissionais e estudantes, bem como na promoção da transparência, ética e responsabilidade social no uso das tecnologias computacionais em todas as áreas da sociedade.

Apesar das conjunturas por vezes desfavoráveis, a SBC continua a avançar, com o surgimento de novas comissões especiais, grupos de interesse e eventos temáticos, e uma crescente interlocução com outros setores da sociedade. Ações estratégicas em áreas como mudanças climáticas, educação básica e promoção da ciência aberta têm sido promovidas, juntamente com a criação de espaços de discussão e difusão de questões emergentes, como os desafios da relação físi-



RAIMUNDO JOSÉ DE ARAÚJO MACÊDO

Presidente da Sociedade Brasileira de Computação (SBC)

co-digital, tema do congresso anual de 2023 em João Pessoa, e o recente painel sobre assistentes virtuais inteligentes.

A revista *Computação Brasil* tem sido mais um espaço para tratar desses temas emergentes, sendo este número dedicado à Mineração de Processos, uma técnica de análise de dados cada vez mais importante na área de computação. Agradecemos ao editor desta revista e sua equipe, bem como aos editores convidados deste número, em particular.

Esperamos que nossos associados e associadas possam aproveitar as valiosas informações disponibilizadas aqui para o contínuo desenvolvimento da computação e suas tecnologias, com o objetivo de contribuir para o desenvolvimento social e econômico do Brasil.

Saudações,

Raimundo José de Araújo Macêdo

abril/2023 • n. 49

COMPUTAÇÃO[®]

REVISTA DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO — BRASIL

Caixa Postal 15012
CEP: 91.501-970 – Porto Alegre/RS
Av. Bento Gonçalves, 9.500 - Setor 4 – Prédio 43412 – Sala 219
Bairro Agronomia - CEP: 91.509-900 - Porto Alegre/RS
Fone: (51) 3308.6835 | Fax: (51) 3308.7142
marketing@sbc.org.br | sbc.org.br

Diretoria:

Presidente | Raimundo José de Araújo Macêdo (UFBA)
Vice-Presidente | André Carlos Ponce de Leon Ferreira de Carvalho (USP)
Diretora Administrativa | Renata Galante (UFRGS)
Diretor de Finanças | Carlos Ferraz (UFPE)
Diretor de Eventos e Comissões Especiais | Cristiano Maciel (UFMT)
Diretora de Educação | Itana Maria de Souza Gimenes (UEM)
Diretor de Publicações | José Viterbo Filho (UFF)
Diretora de Planejamento e Programas Especiais | Tanara Lauschner (UFAM)
Diretor de Secretarias Regionais | Marcelo Duduchi (CEETEPS)
Diretor de Divulgação e Marketing | Alirio Santos Sá (UFBA)
Diretor de Relações Profissionais | Jair Cavalcanti Leite (UFRN)
Diretor de Competições Científicas | Carlos Eduardo Ferreira (USP)
Diretor de Cooperação com Sociedades Científicas | Wagner Meira (UFMG)
Diretora de Articulação de Empresas | Michelle Wangham (UNIVALI)
Diretora de Ensino de Computação na Educação Básica | Leila Ribeiro (UFRGS)

Editor Responsável | Alirio Sá (UFBA)

Editores Convidados | Sarajane Peres (USP), Marcelo Fantinato (USP) e
Eduardo Alves Portela Santos (UFPR)

Equipe de Marketing | Caroline Bittencourt, Cris Felix e Wangles Oliveira

Os artigos publicados nesta edição são de responsabilidade dos autores e não representam necessariamente a opinião da SBC.

Diagramação: Priscila Krüger | priscilahbk@gmail.com | 84 99112-7473

Revisão: Carla Simões de Azevedo

Imagens Ilustrativas: Unsplash.com





ACESSE A SBC OPENLIB

A SBC OpenLib concentra o maior acervo de produções científicas em Computação no Brasil. São mais de 16,8mil itens publicados, entre anais de eventos, periódicos e livros, com acesso aberto e universal.



ACESSE A SOL: [SOL.SBC.ORG.BR](https://sol.sbc.org.br)

ÍNDICE

Mineração de Processos: do que se trata? E o Brasil, está no jogo?

Computação Brasil | Abril 2023

02

EDITORIAL

Raimundo José de Araújo Macêdo

06

MINERAÇÃO DE PROCESSOS: DO QUE SE TRATA? E O BRASIL, ESTÁ NO JOGO?

Sarajane M. Peres, Marcelo Fantinato e Eduardo Alves Portela Santos

10

MINERAÇÃO DE PROCESSOS: DESCOBRINDO O MODELO DO PROCESSO E VERIFICANDO A CONFORMIDADE

16

MINERAÇÃO DE PROCESSOS PARA MONITORAMENTO PREDITIVO

20

MINERAÇÃO DE PROCESSOS E APRENDIZADO DE MÁQUINA: CONQUISTAS DISTRIBUÍDAS, MAS DESAFIOS COMPARTILHADOS



A mineração de processos, enquanto disciplina, permite explicar de forma sistemática fenômenos associados a processos.

-Sarajane M. Peres, Marcelo Fantinato, Eduardo Alves Portela Santos, p. 06

25

MINERAÇÃO DE PROCESSOS NO APOIO À SAÚDE: PERSPECTIVA CLÍNICA E DE GESTÃO

30

MINERAÇÃO DE PROCESSOS APLICADA A PRODUÇÃO E LOGÍSTICA: DESAFIOS PARA A ADOÇÃO DA TÉCNICA PELA INDÚSTRIA

34

MINERAÇÃO DE PROCESSOS: OPORTUNIDADES PARA A TRANSPARÊNCIA PÚBLICA

39

MINERAÇÃO DE PROCESSOS NO JUDICIÁRIO BRASILEIRO

44

MINERAÇÃO DE PROCESSOS EDUCACIONAIS



APRESENTAÇÃO

MINERAÇÃO DE PROCESSOS: DO QUE SE TRATA? E O BRASIL, ESTÁ NO JOGO?

POR

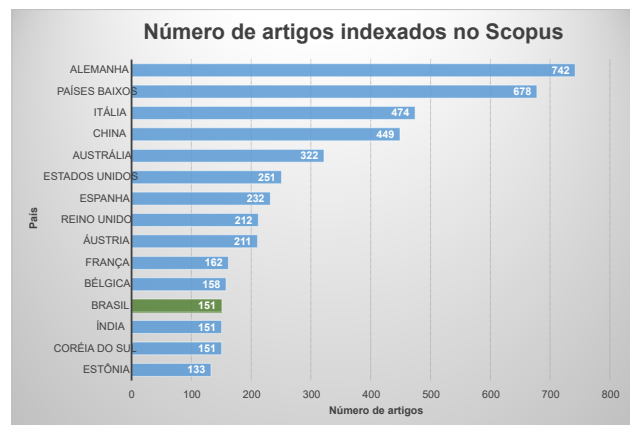
Sarajane M. Peres, Marcelo Fantinato, Eduardo Alves Portela Santos
sarajane@usp.br, m.fantinato@usp.br e portela@ufpr.br

Historicamente, a gestão de processos de negócio tem sido realizada com o apoio de ferramentas computacionais para tarefas tais como modelagem, simulação, e execução e monitoramento de processos. Entretanto, a partir da proposição da **mineração de processos** pelo pesquisador Wil van der Aalst, na virada dos anos 2000, essa gestão passou a contar com o uso de dados referentes à execução de processos de negócio para apoiar auto-

maticamente as tarefas de descoberta de modelo de processo, de verificação de conformidade e de monitoramento e otimização (melhoria) de processos -- essas atividades são ditas “tipos de mineração de processos”. Tais dados dizem respeito, na realidade, ao registro de informações associadas às execuções das atividades pertinentes à lógica de trabalho do processo, ou seja, aos eventos que ocorrem durante o tempo de execução do processo e que são ordenadamente registrados em um **log de eventos**. A mineração de processos, enquanto disciplina, permite

explicar de forma sistemática fenômenos associados a processos. Enquanto fornecedora de ferramentas teóricas e práticas, ela traz assertividade, robustez e eficiência ao trabalho de gestão de processos; já enquanto ramo do conhecimento, promove e impõe um compromisso entre ciência de dados e ciência de processos. As informações do log de eventos podem ser exploradas de diferentes maneiras, havendo espaço para análises preditivas e prescritivas, além da ênfase descritiva. Assim, evidencia-se um leque de oportunidades para gerenciamento e monitoramento do processo e para o desenvolvimento de procedimentos de melhoria. Existem oportunidades para o trabalho apoiado por diversas áreas da análise de dados e da otimização e para a transferência da ‘mentalidade process mining’ para diversas áreas de aplicação.

Para posicionar o Brasil na pesquisa mundial em mineração de processos, consideremos uma busca por artigos científicos realizada na ferramenta Scopus (Elsevier). A busca foi realizada em janeiro de 2023 usando a palavra chave “*process mining*” nos campos “*title, abstract, keywords*” e com uso de filtro de país de afiliação dos autores. A figura 2 mostra os 15 países mais frequentemente encontrados nas afiliações dos autores (a) e o número de artigos publicados por ano por autores afiliados a instituições brasileiras (b). O Brasil está na décima segunda posição em número de publicações, junto com Índia e Coreia do Sul. Na figura 2b, é perceptível um crescimento no número de artigos publicados por brasileiros a partir de 2017.



(a)



(b)



FIG. 01 | NÚMERO DE ARTIGOS PUBLICADOS NA ÁREA DE MINERAÇÃO DE PROCESSOS POR PESQUISADORES COM AFILIAÇÃO BRASILEIRA. BASE DE DADOS: SCOPUS. BUSCA REALIZADA EM JANEIRO DE 2023. A QUEDA DE PUBLICAÇÕES EM 2022 PODE SER, PELO MENOS EM PARTE, RESULTANTE DO ATRASO NATURAL DE INDEXAÇÃO NA BASE DE DADOS.

Em termos da indústria, o uso de mineração de processos está aquém do que já seria possível e desejável¹. Na Europa, o uso de mineração de processos é mais frequente do que em economias importantes como Estados Unidos ou China. Por aqui, pesquisas indicam que há potencial para crescimento. Em 2020, a consultoria HSPI² levantou 500 projetos envolvendo a prática de mineração de processos no mundo. A Europa concentrava 37,9% dos projetos na época. O Brasil concentrava 4,0%, atrás apenas dos Estados Unidos

1 Com a contribuição do Dr. Leonardo Melo Lins, pesquisador no NIC.br (<https://nic.br/>).

2 <https://www.hspi.it/>

(5,1%) (veja o relatório na página da IEEE Task Force on Process Mining³). Uma pesquisa executada pela Deloitte Center for Process Bionics⁴ prospectou em 2021 como as organizações estavam adotando mineração de processos. A pesquisa foi conduzida com 106 empresas das quais 14% eram brasileiras - o terceiro país mais presente na pesquisa. Recentemente, a consultoria ISG⁵ fez um levantamento sobre o movimento da indústria brasileira na adoção de plataformas de automação de ponta a ponta para suas operações. Nesse contexto, a mineração de processos tem sido vista como um passo essencial, porém, segundo a consultoria, as empresas brasileiras enfrentam dificuldades de implantação devido ao alto custo e a preocupações com relação ao uso dos dados necessários. Mineração de processos lida com questões sensíveis e estratégicas para a organização e exige cuidado com privacidade e segurança dos dados básicos (logs de eventos) e demanda recursos humanos altamente capacitados.

Nesta edição da Computação Brasil, pesquisadores brasileiros trazem um resumo da área em termos de definições, oportunidades de pesquisa de base e aplicada. Esses autores representam um conjunto crescente de pesquisadores brasileiros que vem trabalhando com mineração de processos na pesquisa e na extensão universitária, muitas vezes em colaboração com empresas de desenvolvimento

3 <https://www.tf-pm.org/upload/1579851074396.pdf>

4 <https://www2.deloitte.com/de/de/pages/finance/articles/global-process-mining-survey-2021.html>

5 <https://research.isg-one.com/reportaction/Quadrant-AI-PP-Brazil-2022/Marketing>

e consultoria especializada. A apresentação dessa área no contexto da revista tem o intuito de motivar o envolvimento de mais pessoas com mineração de processos, tanto na pesquisa acadêmica quanto na aplicação na indústria.

Os tipos de mineração de processos primeiramente desenvolvidos e mais comumente aplicados são a descoberta do modelo de processo e a verificação de conformidade. Um contexto de utilidade para esses dois tipos de mineração de processos promove o entendimento básico sobre o objetivo da área e é assunto do primeiro artigo desta edição. Conceitos básicos da área, úteis para o acompanhamento do conteúdo aqui apresentado, são fornecidos na forma de um infográfico. O restante desta edição apresenta ao leitor algumas oportunidades de desenvolvimento da área, com destaque para aquelas onde está o principal esforço de pesquisa nacional.

A contextualização da análise preditiva e prescritiva dentro dos objetivos de mineração de processos é apresentada sob duas perspectivas no segundo e terceiro artigos desta edição. A perspectiva do monitoramento de processos é apresentada no segundo artigo por Denise Maria Vecino Sato, Deborah Ribeiro Carvalho e Edson Emílio Scalabrin. Os autores exploram as etapas de monitoramento, a sua importância como apoio operacional para a gestão de processos e fazem um importante apontamento para o caráter dinâmico dos processos. No terceiro artigo, por sua vez, a possibilidade de trabalhar com aprendizado de máquina em mineração de processos é trazida por

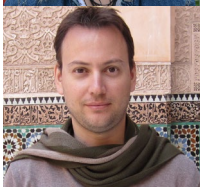
Sylvio Barbon Junior e Sarajane Marques Peres, na visão tanto da análise preditiva quanto da análise prescritiva.

Os cinco artigos seguintes formam um retrato da pesquisa aplicada que tem sido desenvolvida no Brasil. Márcia Ito, Deborah Ribeiro Carvalho e Claudia Moro abordam, no quarto artigo, uma das principais áreas de aplicação de mineração do processo, no Brasil e no mundo: a área de gestão em saúde. As autoras destacam a contribuição que a aplicação na área da saúde proporcionou para a mineração de processos em geral ao apresentar um contexto de gestão de alta variabilidade e complexidade. O quinto, sexto e sétimo artigos mostram a mineração de processos aplicada de forma inovadora. No quinto artigo, Eduardo Alves Portela Santos, Silvana Pereira Detro e Alexandre Choueri Checoli defendem que é possível e adequado levar a mineração de processos, originalmente pensada para a gestão em áreas de negócio, para a gestão de produção e de logística ou para a gestão em ambientes industriais. No sexto artigo, uma oportunidade que foi recentemente

explorada no Brasil, e que ainda não tem sido desenvolvida como poderia, mesmo em âmbito internacional, é apresentada por Renata Araujo e Marcelo Fantinato. Trata-se do uso da mineração de processos como ferramenta para promoção da transparência pública e que tem um grande potencial de impacto para o exercício da democracia. Já o sétimo artigo, de autoria de Ricardo Massa F. Lima, Raphael J. D'Castro, Adriano L. I. Oliveira, Rafael L. Paulo, Bráulio G. Gusmão e João Thiago de F. Guerra, também traz uma área de aplicação que tem sido principalmente desenvolvida no Brasil: o judiciário. Os autores mostram a adequabilidade de usar mineração de processos na gestão de processos jurídicos e apresentam a ferramenta JuMP (Judiciário com Mineração de Processos), de desenvolvimento nacional. Por fim, Petrônio Cândido de Lima mostra, no oitavo artigo, que a mineração de processos pode ser uma importante aliada na melhoria dos processos educacionais, adicionando a perspectiva do processo como uma importante fonte de informação para a análise da aprendizagem.



SARAJANE MARQUES PERES é professora associada na Universidade de São Paulo. Doutora em Engenharia Elétrica pela Universidade Estadual de Campinas e Livre-docente em Inteligência Computacional e Mineração de Dados pela Universidade de São Paulo. Atua em inteligência artificial e mineração de processos. Foi pesquisadora visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



MARCELO FANTINATO é professor associado na Universidade de São Paulo. Doutor em Ciência da Computação pela Universidade Estadual de Campinas e Livre-docente em Gestão de Processos de Negócio pela Universidade de São Paulo. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



EDUARDO ALVES PORTELA SANTOS é professor adjunto na Universidade Federal do Paraná. Doutor em Engenharia Elétrica pela Universidade Federal de Santa Catarina. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Eindhoven University of Technology, nos Países Baixos.



ARTIGO

MINERAÇÃO DE PROCESSOS: DESCOBRINDO O MODELO DO PROCESSO E VERIFICANDO A CONFORMIDADE

POR

Sarajane M. Peres, Marcelo Fantinato, Eduardo Alves Portela Santos
sarajane@usp.br, m.fantinato@usp.br e portela@ufpr.br

Imagine que você é responsável por um processo cuja execução visa alcançar um objetivo de negócio de uma organização. Uma de suas responsabilidades é garantir, no dia a dia da organização, que as atividades desse processo sejam executadas como esperado, ou seja, de forma eficiente e eficaz, e envolvendo os recursos humanos e técnicos corretos. Se esse é seu papel, então você é um gestor de processo de negócio e comumente possui várias perguntas cujas respostas são essenciais para que seu trabalho surta o efeito desejado. Por exemplo, considerando um processo de atendimento ao cliente, as

seguintes perguntas podem surgir:

- Quais são as situações que ocorrem nos atendimentos a domicílio que mais frequentemente levam à necessidade de visitas adicionais ao cliente?
- Sabendo que há atendimentos com substituição de peças que possuem um número de reclamações além do esperado, o que pode ser a causa da insatisfação em relação a esse tipo de atendimento?
- Existem desvios no procedimento esperado para atendimento a clientes? Esses desvios são aceitáveis ou precisam ser corrigidos?

Observando a primeira pergunta com uma lupa, você perceberá que a gestão de processos precisa do apoio de ferramentas que lhe informem sobre a dinâmica do trabalho que está sendo realizado. Uma ferramenta útil é o modelo do processo referente ao atendimento ao cliente. Quando disponível, um modelo de processo representa a lógica de trabalho associada às execuções daquele processo, por exemplo, quais ações são executadas nos atendimentos a clientes e em que ordem elas ocorrem. Se construído e explorado apropriadamente, o modelo de processo pode, inclusive, ajudar a responder todas as perguntas elencadas acima.

A construção de um modelo de processo é parte do ciclo de vida da gestão de processos de negócio. Com uma representação explícita do processo de negócio, podemos sujeitá-lo a análises e melhorias [6]. Além de formalizar a lógica esperada (na visão normativa) ou observada (na visão descritiva), o modelo de processo aliado a procedimentos da gestão e a dados obtidos nos ambientes de execução do processo apoiam ações de alinhamento de negócio, auditoria, redesenho, monitoramento e de otimização de processo.

A mineração de processos permite que dados sobre a execução do processo possam ser usados para automaticamente formalizar a lógica seguida, isto é, descobrir um modelo do processo, dito *descritivo*, a partir do que de fato ocorre no ambiente daquele negócio. Nesse caso, os dados dizem respeito ao registro das atividades que foram executadas, acompanhadas de informações como dia e hora

de execução, tempo de duração, recurso usado para sua execução, custo etc. Esses dados são registrados nos chamados log de eventos (Figura 1).

Case ID	Activity	Start Timestamp	Complete Timestamp
60	Registrar atendimento	2022-08-02 11:00:00.000	2022-08-02 11:10:00.000
60	Agendar visita	2022-08-11 10:13:32.160	2022-08-11 12:00:37.836
60	Realizar visita	2023-04-04 06:46:21.470	2023-04-11 09:56:44.476
60	Cancelar atendimento	2023-10-25 11:35:12.046	2023-10-26 08:22:30.136
60	Fechar atendimento	2023-11-27 13:11:02.533	2023-11-27 13:11:03.533
61	Registrar atendimento	2022-08-02 11:10:00.000	2022-08-02 11:20:00.000
61	Resolver problema imediatamente	2022-08-11 12:00:37.836	2022-08-11 12:07:43.042
61	Registrar resolução do atendimento	2023-04-11 09:56:44.476	2023-04-11 10:26:44.476
61	Cobrar atendimento	2023-10-26 08:22:30.136	2023-10-26 08:52:30.136
61	Fechar atendimento	2023-11-27 13:11:03.533	2023-11-27 13:11:04.533



FIG. 01 | LOG DE EVENTOS REFERENTE A UM PROCESSO DE ATENDIMENTO AO CLIENTE, COM INFORMAÇÕES BÁSICAS PARA A EXECUÇÃO DE UM ALGORITMO DE DESCOBERTA DE MODELO DE PROCESSO.

Vamos clarear essa discussão considerando o log de eventos apresentado na Figura 1. O resultado de um algoritmo de descoberta de modelo é mostrado na Figura 2a (em notação BPMN¹) e nos dá a visão geral do fluxo de trabalho realizado no processo segundo o que foi registrado no log de eventos. Os detalhes na Figura 2b mostram que, nos atendimentos a domicílio, as situações que causam mais de uma visita ao cliente são duas: a ausência do cliente no momento da visita e a necessidade de solicitar uma peça para substituição para resolver o problema. Pelos valores apresentados no modelo, vemos que a segunda situação ocorre com mais frequência. Ainda, quando olhamos o modelo com informação sobre duração das atividades (Figura 2c), percebemos que a solicitação de uma peça demora em média 5,6 dias, enquanto o registro sobre a ausência do cliente demora em média 2,91 horas. Embora essas sejam apenas algumas informações que podem ser extraídas da análise do modelo de processo descoberto automaticamente, elas

1 BPMN - Business Process Model and Notation.

potencialmente esclarecem por que os clientes apresentam mais reclamações quando o atendimento envolve substituição de peças, já que provavelmente o tempo de espera médio entre uma visita e outra pode não estar dentro do esperado do ponto de vista do cliente.

Finalmente, a verificação de conformidade permite responder a terceira

pergunta. Nesse tipo de mineração de processos, um modelo de processo normativo é contrastado ao log de eventos para que seja verificado se a execução do processo difere da lógica considerada como esperada. Em nosso exemplo, um possível desvio no atendimento ao cliente pode ser observado no detalhe ilustrado na Figura 2d. Vamos supor que o modelo

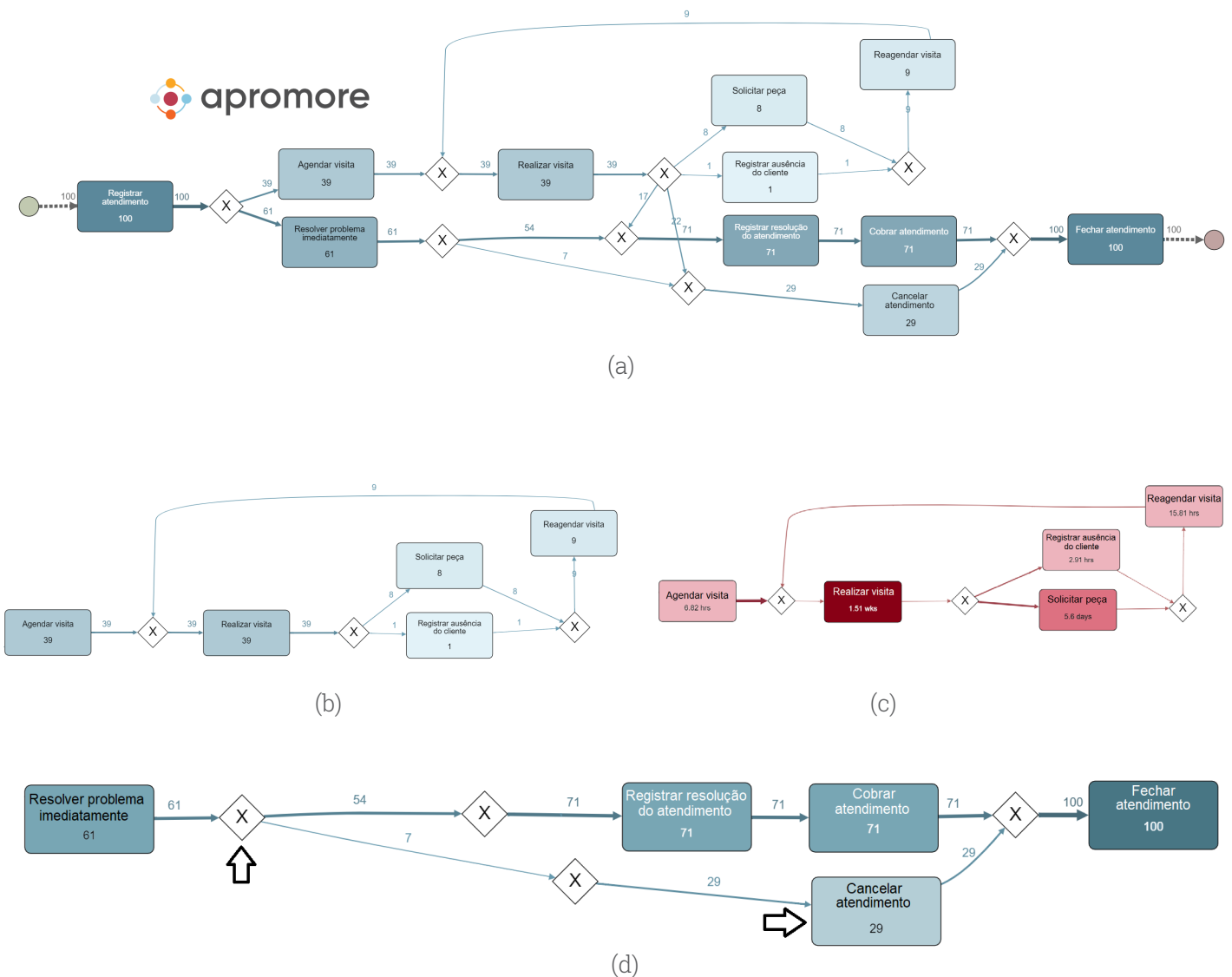


FIG. 02 | MODELO DE PROCESSO DESCOBERTO AUTOMATICAMENTE A PARTIR DO LOG DE EVENTOS. NOS MODELOS DE PROCESSO, RETÂNGULOS COM CANTOS ARREDONDADOS REPRESENTAM ATIVIDADES DO PROCESSO, FLECHAS INDICAM O CAMINHO NO FLUXO DE TRABALHO E LOSANGOS INDICAM A NECESSIDADE DE UMA ESCOLHA ENTRE OS CAMINHOS POSSÍVEIS. OS TONS DAS CORES AZUL E VERMELHO INDICAM FREQUÊNCIA E TEMPO, RESPECTIVAMENTE. QUANTO MAIS INTENSO O TOM, MAIS FREQUENTE OU MAIS DEMORADA É A ATIVIDADE. A ESPESSURA DAS FLECHAS POSSUI SIGNIFICADO EQUIVALENTE. OS VALORES NUMÉRICOS TAMBÉM REPRESENTAM FREQUÊNCIA E TEMPO NAS SITUAÇÕES APROPRIADAS. MODELO DESCOBERTO USANDO A FERRAMENTA APROMORE (APROMORE.COM/ | GRÁFICOS EDITADOS PARA MELHORIA DA VISUALIZAÇÃO).

normativo para esse processo de negócio indique que, uma vez que um atendimento é resolvido, uma cobrança deve ser realizada, seguida pelo fechamento do atendimento. Se confrontarmos essa regra esperada com o registro presente no log de eventos e revelado pelo modelo descritivo (descoberto), veremos que uma violação à regra tem ocorrido, pois há a possibilidade de cancelamento depois da resolução imediata do atendimento ser realizada. Uma vez observada essa violação à norma, o gestor do processo precisa analisar se medidas corretivas precisam ser tomadas ou se a norma precisa ser revista em função da revelação e, talvez adequação, da prática observada.

A descoberta automatizada de modelos de processo e a verificação de conformidade já colocam a mineração de processos como um agente importante para a gestão eficiente e eficaz de processos de negócio. Porém, a mineração de processos vai além dessas tarefas, propiciando inúmeras oportunidades de melhoria e otimização de processos. Um estudo detalhado da área é necessário para que seja possível entender todo o seu potencial. Como uma forma de embasar iniciativas de aprofundamento da área, finalizamos esse artigo fornecendo um infográfico com apontamentos para conceitos, algoritmos, métodos e ferramentas que constituem um mapa para direcionar aqueles que pretendem aprender um pouco mais.

Referências

1. Augusto, A., Conforti, R., Marlon, M., La Rosa, M. & Polyvyanyy, A.: Split Miner: Automated Discovery of Accurate and Simple Business Process Models from Event Logs. *Knowledge and Information Systems*, 59(2), pp. 251–284, 2019.
2. Günther, C.W. & van der Aalst, W. Fuzzy Mining: Adaptive Process Simplification Based on Multi-perspective Metrics. In *International Conference on Business Process Management (BPM 2007)*, v. 4714 of LNCS, Springer, Berlin, pp. 328–343, 2007.
3. Fantinato, M., Peres, S. M. Reijers, H. A. X-Processes: Discovering More Accurate Business Process Models with a Genetic Algorithms Method. In: *25th IEEE International Enterprise Distributed Object Computing Conference*, (online) Gold Cost, p. 114–123, 2021.
4. Leemans, S.J.J., Fahland, D. & van der Aalst, W. Discovering Block-Structured Process Models from Event Logs Containing Infrequent Behaviour. In *Business Process Management Workshops (BPI 2013)*, v. 171 of LNBIP, Springer, Berlin, pp. 66–78, 2014.
5. van der Aalst, W. *Process Mining - Data Science in Action*. 2nd ed. Springer, 2016.
6. Weske, M. *Business Process Management - Concepts, Languages, Architectures*. 2nd ed. Springer, 2012.



SARAJANE MARQUES PERES é professora associada na Universidade de São Paulo. Doutora em Engenharia Elétrica pela Universidade Estadual de Campinas e Livre-docente em Inteligência Computacional e Mineração de Dados pela Universidade de São Paulo. Atua em inteligência artificial e mineração de processos. Foi pesquisadora visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



MARCELO FANTINATO é professor associado na Universidade de São Paulo. Doutor em Ciência da Computação pela Universidade Estadual de Campinas e Livre-docente em Gestão de Processos de Negócio pela Universidade de São Paulo. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



EDUARDO ALVES PORTELA SANTOS é professor adjunto na Universidade Federal do Paraná. Doutor em Engenharia Elétrica pela Universidade Federal de Santa Catarina. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Eindhoven University of Technology, nos Países Baixos.



Conceitos básicos

Um **evento** é uma observação da ocorrência de um fenômeno associado a um processo de negócio, geralmente produzido por um recurso (uma pessoa) com o apoio de um sistema de informação – como ocorrência de um fenômeno, o evento é único no tempo e no espaço.

e

$\#_{att}(e)$

A cada evento associa-se um conjunto de **atributos** que o caracteriza, por exemplo, atividade, *timestamp*, recurso e custo. No contexto de um processo, o evento está associado a uma atividade, ocorre em um determinado momento do tempo, a partir da ação de um recurso (pessoa, sistema, máquina) e pode ter um custo associado.

Para fins de praticidade, é comum mapearmos o evento para o conteúdo de seus atributos, ou seja, no decorrer do desenvolvimento de uma solução em mineração de processo, podemos, por exemplo, mapear o evento diretamente para o valor da atividade cuja execução o originou – dizemos que a função que faz esse mapeamento é um **classificador**.

e

C Um **caso** diz respeito a uma instância de processo, ou seja, uma execução do processo. O caso é formado por uma sequência de eventos únicos e também possui atributos: um *mandatário*, denominado *trace*.

O *trace* diz respeito especificamente à sequência de eventos, ordenada no tempo, que compõem o caso.

σ

L

O **log de eventos** é um conjunto de casos.



- Um **log de eventos simples** é obtido ao aplicar um classificador sobre os eventos, mapeando-o para, por exemplo, os valores do atributo *atividade*.

- O atributo *trace* também é mapeado dando origem ao **trace simples**: uma sequência de valores de atributo. Por consequência, o log de eventos simples é um multiconjunto de *traces* simples.

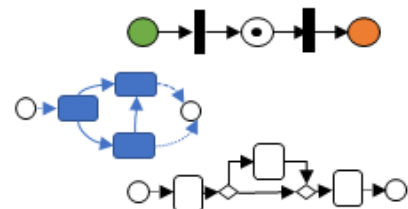


Na prática de mineração de processos, é bastante comum encontramos a denominação **variante** para uma determinada sequência de atividades existente no log de eventos. De fato, uma *trace* simples diz respeito a uma variante. Ainda, por vezes, usa-se a denominação *trace* para indicar o *trace* simples, ou a variante.



Notações para representação mais comuns

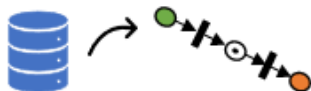
- Petri Net** (e uma variação especial chamada *Workflow-net*).
- DFG** – *Direct Flow Graph*.
- BPMN** – *Business Process Model Notation*.



Relacionamento entre o modelo de processo e a realidade capturada pelo log de eventos

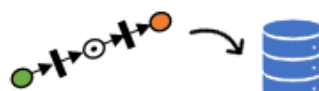
play in

O comportamento é entrada para construção do modelo!



play out

Dado um modelo, gera-se o comportamento!



replay

O log de eventos é reproduzido sobre um modelo!





Conceitos básicos

Critérios de qualidade

Fitness (ou revocação): avalia se um modelo gera o comportamento presente no log de eventos. Um modelo alcança o valor máximo nesse critério se todos os *traces* presentes no log de eventos podem ser reproduzidos, do início ao fim, no modelo de processo.



Precisão: avalia se um modelo de processo não permite comportamentos muito diferentes daqueles presentes no log de eventos. Um modelo é totalmente preciso se nenhum *trace* completamente diferente dos *traces* registrados no log de eventos pode ser produzido pelo modelo. Por um lado, um modelo totalmente preciso pode ser considerado um modelo **super ajustado** ao log de eventos; por outro lado, baixos valores para a precisão podem indicar o **sub ajuste** do modelo.

Generalização: avalia se o modelo de processo não está super ajustado, ou seja, se o modelo permite a geração de comportamento não presente no log de eventos, porém similar ou relacionado aos comportamentos registrados.



Simplicidade: único critério cuja avaliação é extrínseca aos comportamentos registrados no log, ele avalia se o modelo de processo é tão simples quanto possível. Em geral, quanto menos elementos forem usados, mais simples será o modelo.

Principais algoritmos para descoberta

- **Algoritmo Alpha** (α -algorithm): primeiro algoritmo desenvolvido para descoberta de modelos de processo por van der Aalst [5].
- **Fuzzy Miner:** algoritmo apresentado por Günther e van der Aalst [2], se tornou popular para a implementação de ferramentas industriais.
- **Inductive Miner:** baseado em uma abordagem de dividir-para-conquistar, foi apresentado por Leemans, Fahland e Aalst [4].
- **Split Miner:** considerando atualmente o estado da arte dos métodos de descoberta de processos, esse algoritmo foi apresentado por Augusto *et al.* [1].



O algoritmo X-Processes, baseado em algoritmos genético, foi desenvolvido pelo grupo **Process Mining @ USP**, da **Universidade de São Paulo**. O algoritmo tem apresentado resultados promissores em termos de qualidade de modelo de processo descoberto considerando a otimização do balanceamento dos critérios de qualidade comumente usados em mineração de processos. [3]



Token replay

Estratégias para checagem de conformidade



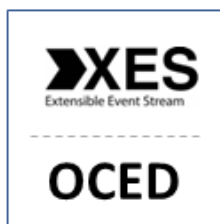
Alinhamento

Padrões para log de eventos

Exemplos de Ferramentas open-source

Exemplos de Ferramentas industriais

(com programa de licença acadêmica ou cursos associados)





ARTIGO

MINERAÇÃO DE PROCESSOS PARA MONITORAMENTO PREDITIVO

POR

Denise Maria Vecino Sato, Deborah Ribeiro Carvalho e Edson Emílio Scalabrin
denise.sato@ifpr.edu.br, ribeiro.carvalho@pucpr.br e scalabrin@ppgia.pucpr.br

A área de mineração de processos disponibiliza diferentes técnicas e ferramentas para monitorar os processos de negócio, permitindo visualizar como eles ocorrem na prática a partir dos dados armazenados pelos sistemas de informação [5]. Você tem interesse em conhecer seus processos como realmente acontecem no dia a dia? Essas técnicas e ferramentas com certeza podem lhe ajudar nisso, pois, através dos dados obtidos em logs de eventos é possível visualizar o

modelo do seu processo (descoberta), compará-lo com o que deveria acontecer (verificação de conformidade) ou enriquecê-lo com mais dados que estejam disponíveis para uma visão mais ampla (enriquecimento). Inicialmente, você pode obter um log de eventos de um período passado, contendo *traces* completos e finalizados, e então aplicar técnicas que permitem conhecer e entender o que já ocorreu no seu processo. A partir dessa análise do passado, utilizando a mineração de processos, é possível traçar planos de

melhoria. No entanto, podemos ir além do entendimento e análise do que já passou, aproveitando o conhecimento obtido para monitorar o que está acontecendo e atuar nas situações em andamento. Se conhecer o que ocorreu já é um grande avanço para identificação de problemas e melhorias em situações futuras, imagine se você puder combinar esse conhecimento e atuar nas situações ainda em andamento? O termo suporte operacional é utilizado para indicar a combinação de dados já ocorridos no processo, utilizando log de eventos com *traces* completos, com dados de situações em andamento, descritos por *traces* parciais [5]. Através do suporte operacional é possível atuar nos casos em curso, influenciando nos seus andamentos e desfechos, por exemplo, prevendo que a entrega do seu produto para o cliente vai atrasar ou mesmo identificando a possibilidade de falta de algum recurso com antecedência. Uma forma de atuar sobre os casos em andamento é utilizando o *Monitoramento Preditivo de Processos*, que permite prever informações futuras sobre casos ainda em andamento, ou seja, incompletos [2]. Vamos entender melhor o que poderia ser essa predição em uma situação prática. .

Podemos considerar o cenário de um paciente que é atendido pelo Serviço de Urgência e Emergência, percorrendo o processo que acompanha todas as atividades desse atendimento: acolhimento pela equipe administrativa, avaliação de enfermagem para classificação de risco (definição da priorização do atendimento), espera pela consulta médica, a consulta médica em si, realização de exames até um dos possíveis desfechos—internação, alta (libe-

ração para retorno à casa) ou mesmo interrupção do cuidado (fuga). Como o monitoramento preditivo poderia ser aplicado nesse processo? Inicialmente precisamos avaliar qual *tipo de predição* seria interessante para o processo. Então, indicar se a sequência de atividades para um paciente em atendimento incluirá exames e internação é considerada uma *predição do próximo evento* (execução de uma atividade). Também poderíamos prever em quanto tempo o paciente realizará o próximo exame; nesse caso teríamos a predição de um valor numérico, portanto uma *predição de medida*. Já prever o desfecho de um paciente entre as categorias internação, alta ou interrupção, seria a *predição de categorias pré-definidas*. É possível agrupar as saídas do monitoramento nesses três tipos de predição: *categorias pré-definidas, medida ou próximo evento* [2].

Além do tipo de predição, precisamos considerar quais e qual tipo de informação possuímos como *entrada* para o treinamento do nosso modelo. Podemos considerar informações relacionadas a sequência de atividades do processo, ou seja, os históricos anteriores dos pacientes (*traces*). Também podem ser incluídos dados relacionados aos eventos, como quando ocorreu o exame (*timestamp*) ou mesmo em qual sala ele foi realizado e qual médico foi o responsável. Em muitas situações possuímos informação não estruturada, como por exemplo, o histórico médico que foi coletado na triagem com o paciente ou ainda informação relacionado ao contexto do processo, como a disponibilidade de um leito para internamento do paciente. Esses dados podem ser necessários dependendo

do tipo de predição escolhida. Por exemplo, para prever quando o paciente será internado é necessário considerar a disponibilidade de leito além do histórico e situação atual do paciente.

Diferentes *abordagens* foram propostas para construção de modelos preditivos em processos, e podemos agrupá-las em duas categorias: *baseada em modelo*, que utiliza um modelo de processo enriquecido para fornecer predições, ou *aprendizagem supervisionada*, que basicamente utiliza modelos preditivos herdados da área de aprendizagem de máquina ou técnicas estatísticas para gerar as predições [2]. A técnica por trás do modelo preditivo vai buscar sempre uma forma de melhorar a acurácia das predições no cenário analisado, ou seja, quão próximo da realidade a saída do modelo preditivo está.

Apesar de ser uma área relativamente nova, diferentes pesquisas vêm sendo apresentadas na área de monitoramento

preditivo de processo, como pode ser verificado nas revisões disponíveis na literatura [1, 3]. Podemos classificar os trabalhos considerando o *tipo de predição*, a *abordagem adotada* e a *entrada* utilizada [2], como detalhamos no exemplo.

Para aplicar o monitoramento preditivo em processos precisamos de duas etapas: a fase de aprendizagem e treinamento, e a fase de execução e predição, conforme apresentado na Figura 1. Na etapa de treinamento, deve-se considerar o tipo de predição desejado, qual a abordagem a ser adotada e quais dados estão disponíveis para entrada. Nesta etapa, a acurácia do modelo deve ser avaliada para buscar predições mais próximas à realidade. Após o modelo ter sido treinado e avaliado, ele pode ser utilizado na etapa de execução. Aqui, um *trace* incompleto (ainda em andamento) é encaminhado ao modelo com as demais informações de entrada necessárias e o resultado da predição é fornecido.

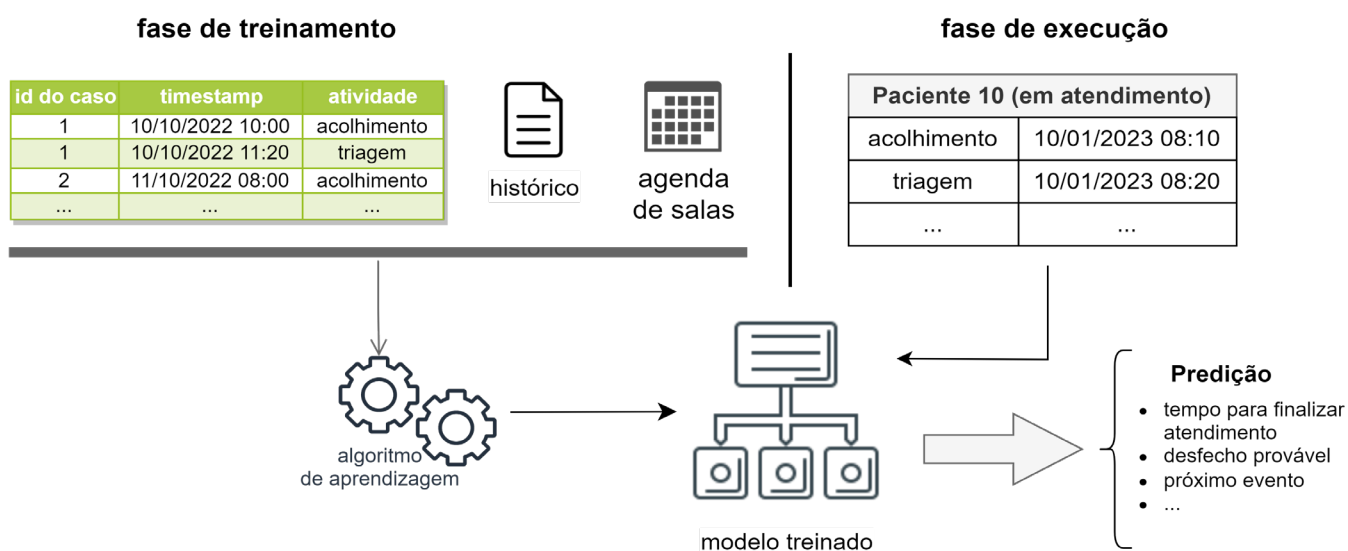


FIG. 01 | ETAPAS DO MONITORAMENTO PREDITIVO DE PROCESSOS.

Apesar de um modelo preditivo ter apresentado uma boa acurácia durante a validação e um período de execução, ele pode se tornar menos preciso ao longo do tempo. Isso muitas vezes está associado ao fato de que os processos são dinâmicos e tendem a mudar e evoluir. A situação

na qual um processo muda em um determinado momento do tempo é chamada de *mudança de conceito* [4]. Os modelos utilizados no monitoramento preditivo de processos devem ser continuamente avaliados para garantir sua acurácia, ou mesmo atualizados em situações de *mudança de conceito*.

Referências

1. DI FRANCESCO MARINO CHIARA AND GHIDINI, C.; M. F. M. AND M. F. Predictive Process Monitoring Methods: Which One Suits Me Best? In: M. and W. I. and vom B. J. Weske Mathias and Montali (Org.); Business Process Management. Anais... p.462–479, 2018. Cham: Springer International Publishing.
2. DI FRANCESCO MARINO CHIARA AND GHIDINI, C. Predictive Process Monitoring. In: J. van der Aalst Wil M. Pand Carmona (Org.); Process Mining Handbook. p.320–346, 2022. Cham: Springer International Publishing.
3. MARQUEZ-CHAMORRO, A. E.; RESINAS, M.; RUIZ-CORTES, A. Predictive monitoring of business processes: A survey. IEEE Transactions on Services Computing, v. 11, n. 6, p. 962–977, 2018.
4. RESINAS, M.; RUIZ-CORT, A.; M, A. E. Predictive Monitoring of Business Processes : A Survey. IEEE Transactions on Services Computing, v. 11, n. 6, p. 962–977, 2018.
5. SATO, D. M. V.; DE FREITAS, S. C.; BARDDAL, J. P.; SCALABRIN, E. E. A Survey on Concept Drift in Process Mining. ACM Computing Surveys, v. 54, n. 9, p. 1–38, 2022. ACM PUB27 New York, NY.
6. VAN DER AALST, W. Operational Support. Process Mining: Data Science in Action. p.301–321, 2016. Berlin, Heidelberg: Springer Berlin Heidelberg.



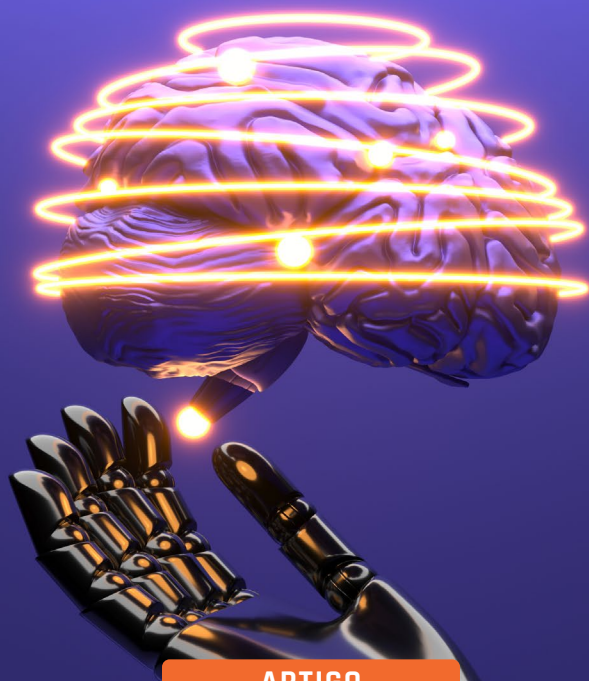
DENISE MARIA VECINO SATO é Professora no Instituto Federal do Paraná - Campus Curitiba, com mestrado (2014) e doutorado em Informática (2022) pela PUCPR. Atua como pesquisadora na área de inteligência artificial, com foco atual na detecção de mudança de conceito na mineração de processos.



DEBORAH RIBEIRO CARVALHO é Professora da Pontifícia Universidade Católica do Paraná - PUCPR, com mestrado (1999) e doutorado (2002) em Informática pela PUCPR e doutorado em Computação de Alto Desempenho pela UFRJ (2005). Atua como pesquisadora em inteligência artificial aplicada à saúde nas áreas clínica e de gestão.



EDSON EMÍLIO SCALABRIN é Professor da Pontifícia Universidade Católica do Paraná, com mestrado (1993) e doutorado (1996) em Controle de Sistemas pela Université de Technologie de Compiègne (França). Atua como pesquisador em inteligência artificial e ciência de dados.



ARTIGO

MINERAÇÃO DE PROCESSOS E APRENDIZADO DE MÁQUINA: CONQUISTAS DISTRIBUÍDAS, MAS DESAFIOS COMPARTILHADOS

POR

Sylvio Barbon Junior, Sarajane Marques Peres
sylvio.barbonjunior@units.it e sarajane@usp.br

Atualmente, a aplicação de métodos fundamentados em dados é praticamente onipresente em cenários de resolução de problemas desafiadores. Desafios como a exigência de alto desempenho na identificação de padrões e a extração de significado em formas de modelagem do conhecimento fazem parte de tais cenários e também estão presentes no gerenciamento de processos de negócio. Desta forma, soluções baseadas em métodos de Aprendizado de Máquina podem ajudar na gestão e otimização de processos de negócio e na inovação em soluções para essa área.

O uso de aprendizado de máquina na mineração de processos segue preceitos semelhantes aos seguidos em mineração de dados. Metodologias e práticas representadas pelos chamados fluxo de etapas, ou *pipelines*, para processamento e modelagem de dados usados em mineração de dados são seguidos, com algumas singularidades, em mineração de processos. Por exemplo, considere um *pipeline* clássico em mineração de dados (Figura 1(a)), constituído por quatro procedimentos encadeados com origem em grandes bases de dados: limpeza de dados, pré-processamento dos dados, desenvolvimento de um modelo (descritivo, preditivo ou prescri-

tivo) usando algoritmos de aprendizado de máquina e, finalmente, implantação ou uso do modelo de acordo com os objetivos de resolução de um problema. Esse *pipeline* pode ser aplicado em mineração de processos, contanto que sejam observadas algumas questões. De forma geral, o ponto nevrálgico do *pipeline* aplicado na mineração de processos está na codificação do log de eventos em um formato adequado para uso pelo algoritmo de aprendizado de máquina. A Figura 1(b) destaca este ponto de atenção.

Análises em mineração de processos têm como origem as informações relacionadas aos eventos advindos da execução de um processo de negócio, armazenados no formato de um log de eventos. O log de eventos é estruturado de acordo com a sequência temporal dos eventos, intercalando informação de diferentes instâncias de um mesmo processo. Essa é uma característica importante, que exige cuidado na etapa de pré-processamento pois impõe dificuldades adicionais na forma como o dado bruto será transformado e apresentado para os algoritmos de aprendizado de máquina. Tais dificuldades estão relacionadas a sequencialidade temporal de eventos dentro de suas respectivas instâncias de processo, existência de relações de causalidade representadas por dependências de longo prazo entre eventos, e possíveis falsas relações de causalidade derivadas da relação temporal imediata entre eventos. Em geral, tarefas de mineração de processos estão interessadas em analisar o comportamento do processo, para além de suas características descritivas factuais. Nem sempre a transformação do dado

para uma representação adequada para um algoritmo de aprendizado de máquina será capaz de carregar informação que respeite as características mencionadas, porém, o viés da representação do dado precisa ser considerado no momento da interpretação do modelo de dados obtido na saída de um algoritmo de aprendizado de máquina.

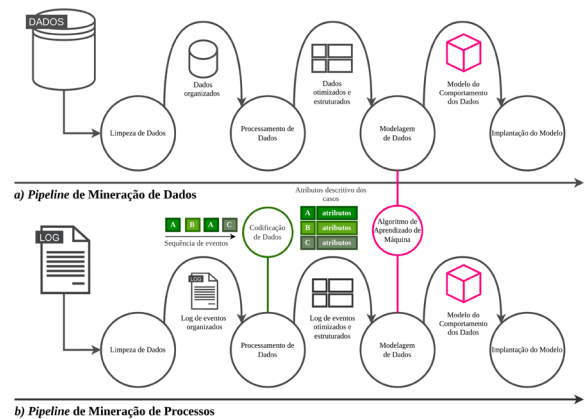


FIG. 01 | COMPARAÇÃO ENTRE OS PIPELINES PARA MINERAÇÃO DE DADOS E MINERAÇÃO DE PROCESSOS

O processo de codificação do log de eventos pode ser realizado por diferentes métodos, cuja principal meta é transformar a informação presente na sequência temporal original de eventos na instância do processo (ou caso) em uma representação estruturada utilizando descritores do comportamento original. Assim como na mineração de Dados, diferentes métodos resultarão em resultados de maior ou menor qualidade, em maior ou menor tempo de processamento e em descrições mais ou menos interpretáveis por especialistas humanos.

A possibilidade de aplicar aprendizado de máquina amplia o espectro de análises e otimizações que podem ser feitas sobre os processos de negócios apoiados pela mineração de Processos. Dentre

as contribuições, é possível desenvolver modelos capazes de extrair padrões referentes ao comportamento do processo para, por mais dinâmico que o processo seja, aprimorar a integridade das atividades nele realizadas, além de acompanhar em tempo de execução como os padrões descobertos evoluem durante a execução de milhares de instâncias do processo. Para ilustrar esse potencial, considere dois exemplos de contribuições atuais em mineração de Processos, construídas com o apoio do aprendizado de máquina: detecção de comportamentos anômalos na execução do processo e predição do tempo, da próxima atividade ou do evento final da execução atual de uma instância do processo.

Detecção de comportamentos anômalos: o comportamento anômalo se caracteriza pelas execuções (instâncias) ou fenômenos que não são esperados por um processo de negócio, sendo essencial para a manutenção da qualidade das operações dentro de uma organização. Exemplos são a detecção de sequências de eventos incorretas e a detecção da ausência da execução de atividades que, se localizadas e corrigidas, podem evitar fraudes e economizar recursos. Tradicionalmente, a detecção de anomalias ocorre durante as tarefas de monitoramento, verificação de conformidade e de otimização. Normalmente, é esperado que a equipe de controle de processos em uma organização seja capaz de detectar comportamentos anômalos para rápida mitigação. No entanto, uma grande quantidade de eventos produzidos em um curto espaço de tempo impede eficiên-

cia e eficácia nesta tarefa. Além disso, a detecção dos mais sensíveis comportamentos em processos complexos exige do colaborador treinamento, experiência e ferramentas especializadas. Considerando o contexto dos processos de negócios atuais, controle completo e em tempo real se tornaram não só uma vantagem competitiva, mas algo essencial para se manter competitivo no mercado. Neste cenário, o aprendizado de máquina surge como capaz de trazer boas soluções. Algoritmos não supervisionados têm sido usados para identificar anomalias *a priori* e então apoiar a construção de modelos supervisionados de monitoramento. Com algoritmos de agrupamento em tempo real (*clustering*), pode-se criar ferramentas que agrupam os comportamentos normais, isolando as instâncias anômalas ainda durante a execução. Uma solução para este cenário foi proposta em Vertuam Neto et al. [3]. Outra alternativa recentemente aplicada na área são os *autoencoders* que, ao comprimir informação mediante extrações sucessivas de características, podem encontrar anomalias referentes a relações estabelecidas à longo prazo, considerando um contexto temporal.

Predição na execução de processos: a riqueza de informação proporcionada pelos logs dos eventos permite a compreensão e a realização de diagnósticos sobre situações subjacentes aos processos de negócio. Além de contribuir para o diagnóstico e mapeamento da situação corrente, os logs de eventos fornecem subsídios para a criação de modelos para previsão (predição) de ações futuras, imple-

mentando o monitoramento preditivo de processos. Exemplos de ações nessa área são a previsão do tempo para encerramento da execução de uma instância, da próxima atividade a ser realizada em uma dada execução e de outras informações sobre o possível desfecho da execução de uma instância do processo. Os algoritmos de aprendizado de máquina, especificamente os supervisionados, são adequados para implementar modelos preditivos. Especificamente para o contexto de mineração de processos, devido à necessidade de processamento de informação sequencial, redes neurais da classe das LSTM têm sido amplamente exploradas. Uma revisão sobre aplicações nas quais o aprendizado de máquina pode ser útil no contexto de previsão pode ser encontrada em Di Francescomarino et al. [1].

Otimização na alocação de recursos:

um ponto que vem sendo cada vez mais crítico em organizações é como gerenciar os seus recursos humanos frente às necessidades das tarefas envolvidas em um processo de negócio. Principalmente por conta do custo de execução das tarefas depender de fatores dinâmicos, a otimização da alocação de recursos se torna uma tarefa complexa que pode ser resolvida com aprendizado de máquina no contexto da mineração de processos. A partir do uso da informação contida no log de eventos, dados históricos referentes a eficiência da alocação de recursos no passado podem ser usados para o estabelecimento de políticas para uma alocação eficiente de recursos no futuro. Como um exemplo da implementação de uma estratégia nesse problema, consulte [2].

Vale ressaltar aqui que problemas como este revelam um lado sensível da mineração de processo: lidar com informações de desempenho humano para otimização de processos pode recair sobre questões éticas e legais. Essa sensibilidade de informação e de seu uso estabelece um campo recente de pesquisa na área conhecida como mineração de processos responsável.

Alinhamento da mineração de processo com o aprendizado de máquina:

Os desafios das organizações com o uso da mineração de Processos vão além da exploração descritiva e do diagnóstico. Soluções criativas e inovadoras para otimização, previsão e prescrição dependem da integração eficiente da mineração de processos com o poder de modelagem do aprendizado de máquina, aqui chamada de alinhamento.

Uma série de desafios precisam ser enfrentados para que a contribuição do alinhamento entre mineração de dados e os algoritmos de aprendizado de máquina seja frutífera. O primeiro deles é a construção de uma representação computacional para a informação contida no log de eventos capaz de evidenciar os padrões que precisam ser aprendidos pelo modelo de aprendizado de máquina. Além disso, cada processo tem suas particularidades e a obtenção de um modelo que reflita adequadamente o comportamento de suas instâncias de execução pode exigir alto custo computacional, fator impeditivo em processos críticos. Desta forma, um desafio bastante atual é encontrar um método de codificação adequado para

o log de eventos a fim de aproveitar com eficácia o poder analítico do aprendizado de máquina no contexto da gestão de processos de negócio.

Outro desafio que merece destaque é a necessidade de lidar com as instâncias do processo incompletas. A maioria dos algoritmos de aprendizado de máquina demanda amostras atômicas para modelar o comportamento de um fenômeno. No entanto, quando uma instância de processo está sendo analisada em aplicações de tempo real, não é possível garantir

que não existirão eventos posteriores ao momento de análise. Esse tipo de situação exige abstrações adicionais durante a modelagem de um problema.

Por fim, também é muito discutido na área de mineração de processos o problema de ausência de exemplos negativos. Naturalmente, a informação com a qual a mineração de processos lida é a informação factual, ou seja, não há informação sobre o que não deve ocorrer. Essa característica impõe restrições particulares para modelagem de problemas.

Referências

1. DI FRANCESCO MARINO, C., GHIDINI, C., MAGGI, F. M., & MILANI, F. (2018). Predictive process monitoring methods: Which one suits me best? In 16th International Conference on Business Process Management, Sydney, NSW, Australia, September 9–14, 2018, (pp. 462-479). Springer International Publishing.
2. NEUBAUER, T. R., da SILVA, V. F., FANTINATO, M., & PERES, S. M. Resource Allocation Optimization in Business Processes Supported by Reinforcement Learning and Process Mining. In 11th Brazilian Conference on Intelligent Systems, Campinas, Brazil, November 28–December 1, 2022, (pp. 580-595). Cham: Springer International Publishing.
3. VERTUAM NETO, R., TAVARES, G., CERAVOLO, P., & BARBON, S. (2021, June). On the use of online clustering for anomaly detection in trace streams. In XVII Brazilian Symposium on Information Systems (pp. 1-8).



SYLVIO BARBON JUNIOR é professor associado na Universidade de Trieste, Itália. Doutor em Física Computacional pela Universidade de São Paulo, atua como professor e pesquisador na área de Reconhecimento de Padrões, com diferentes tipos de dados. Atua em mineração de processos desde 2016; atualmente é um dos coordenadores do workshop em Machine Learning na Mineração de Processos que ocorre anualmente no Congresso Internacional de Mineração de Processos.

SARAJANE MARQUES PERES é professora associada na Universidade de São Paulo. Doutora em Engenharia Elétrica pela Universidade Estadual de Campinas e Livre-docente em Inteligência Computacional e Mineração de Dados pela Universidade de São Paulo. Atua em inteligência artificial e mineração de processos. Foi pesquisadora visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



ARTIGO

MINERAÇÃO DE PROCESSOS NO APOIO À SAÚDE: PERSPECTIVA CLÍNICA E DE GESTÃO

POR

Márcia Ito, Deborah Ribeiro Carvalho e Claudia Moro
marcia.ito01@fatec.sp.gov.br, ribeiro.carvalho@pucbr.br e c.moro@pucpr.br

O planejamento, fornecimento e previsão de recursos, seja num hospital, na atenção primária, ou no sistema de saúde como um todo sempre foi um grande desafio na gestão das organizações e no tratamento de pacientes. Modelos, métodos e processos que possam capturar a complexidade do ecossistema de saúde são necessários. A mineração de processos é uma das técnicas que pode ser utilizada para esta finalidade.

Modelar a trajetória do paciente é considerado vital para entender a operação do sistema como um todo e pode, portanto, se revelar útil para melhorar a qualidade e a gestão desse sistema.

Independente dos atores envolvidos, desde várias instituições (hospitais, indústria farmacêutica, governo, planos e seguradoras de saúde, entre outros) até indivíduos como os profissionais de saúde (médicos, enfermeiras, nutricionistas, psicólogos, etc.), pacientes e cuidadores. Considera-se trajetória ou jornada de paciente a sequência de eventos de saúde ou intervenções que o paciente foi submetido. O evento pode ocorrer numa instituição de saúde ou não, como por exemplo os eventos no *home care*.

Por outro lado, o aumento de casos de doenças crônicas e de pacientes com múltiplas condições (multimorbidade) fazem com que haja uma maior de-

manda por serviços. Neste caso, o planejamento e previsão de recursos vai além da operação do dia a dia das instituições e estão relacionados com a demanda vinda de fatores externos. Um exemplo, foi a pandemia da COVID-19 que comprometeu o planejamento e a previsão, bem como a opção do indivíduo se vacinar ou não, o qual acarretou o aumento na demanda do sistema. Assim, conclui-se que não basta modelar o processo operacional, mas é preciso entender a trajetória do paciente, considerando suas escolhas e modo de vida (comportamento) no planejamento e previsão da gestão de saúde. Neste sentido, a mineração de processos é uma técnica que pode apoiar tanto na modelagem dos processos operacionais quanto na trajetória dos pacientes.

Nas primeiras aplicações de mineração de processos na saúde foram observadas dificuldades relacionadas à representação dos modelos descobertos, provavelmente pelo fato dos algoritmos de mineração de processos terem sido desenvolvidos para a indústria, não adequados à área da saúde. Esta variabilidade está diretamente vinculada a estratégia de cuidado em relação ao paciente, pois é difícil estabelecer a previsibilidade da sua trajetória, que depende de inúmeros fatores desde: interações biológicas, sua patologia, tipo de tratamento realizado, entre outros [2].

É possível acompanhar a trajetória de um único paciente ou conjuntamente de vários e analisar a respectiva conformidade com uma diretriz clínica

padrão e melhorar ou entender o seu processo de tratamento. Isto permite (a) avaliar o desempenho de diferentes unidades no sistema de saúde; (b) encontrar padrões de trajetórias em doenças após eventos críticos; (c) otimizar *layouts* hospitalares; (d) buscar por possíveis atualizações nos protocolos clínicos padrão; (e) analisar o custo e desempenho de procedimentos, tratamentos e diagnósticos; (f) encontrar gargalos nos processos de serviço de saúde, entre outros parâmetros definidos pelo administrador do sistema [4].

As trajetórias de pacientes podem ser observadas sob duas perspectivas:

- gestão do serviço de saúde, que representa a trajetória de atividades organizacionais, departamentos ou outros eventos relacionados ao serviço de saúde;
- clínica, que representa a sequência temporal de procedimentos, diagnósticos e outras atividades relacionadas diretamente ao tratamento, a partir de dados associados ou não aos pacientes tais como registro clínico, sinistro, condições climáticas do dia do evento, entre outros, e não somente de dados clínicos.

Os primeiros estudos de mineração de processos na saúde tiveram como propósito a área de gestão, utilizando dados de acesso menos restritivo, devido à privacidade dos dados clínicos, como os logs de sistemas de informação administrativos de hospitais, operadoras de saúde e dados públicos de vigilância epidemiológica disponíveis

em sistemas governamentais. Nas aplicações relacionadas à área clínica, como acompanhamento de pacientes, diagnóstico e avaliação de desfechos clínicos, também foram utilizados dados de gestão associados aos clínicos e com um olhar mais individualizado para as características e necessidades de saúde de cada indivíduo.

Considerando a adoção de mineração de processos na gestão em saúde pode-se destacar algumas situações, por exemplo, a identificação do itinerário dos usuários para atendimento em Unidades Básicas de Saúde (UBS), o que permite caracterizar a demanda desde a sua entrada na rede até a resolução do motivo que originou a busca (desfecho).

Um estudo usou uma amostra de dados de usuários residentes no município de Curitiba que buscaram por serviços nas UBS e nas unidades de pronto atendimento (UPA), no período de janeiro a dezembro de 2019, disponíveis no Portal de Dados Abertos do Município de Curitiba. Entre os resultados foi identificado que alguns usuários, mesmo vinculados às UBS, procuram por consultas de rotina em UPA em uma janela temporal que contraria o tempo preconizado pela gestão [5], evidenciando a necessidade de um de um olhar mais atento aos motivos por esta inadequação, pois a existência de itinerários que não correspondem ao preconizado sobrecarrega a demanda por atendimentos nas UPAs.

Um outro exemplo remete à gestão de salas de cirurgia em um hospital.

Entre os desafios envolvidos nesta gestão está o atraso na realização de um procedimento cirúrgico que implique em uma sucessão de atrasos ou mesmo no cancelamento de procedimentos que estavam agendados. A partir da mineração de processos é possível prever se um determinado procedimento tende ou não ao atraso, permitindo uma reorganização da agenda.

Na área clínica, a mineração de processos pode ser aplicada no acompanhamento da jornada assistencial para otimizar a experiência dos pacientes e obter os melhores desfechos de saúde, identificando gargalos, ineficiência e possibilidades de melhorias no processo clínico. Uma das aplicações da mineração de processos é no acompanhamento de pacientes com condições crônicas, como o Acidente Vascular Cerebral (AVC), em que a mineração de processos foi utilizada para analisar o atendimento desde os primeiros sintomas até a evolução da doença [3], visando evitar óbitos e recorrência desta condição. Também foram avaliados o impacto da pandemia de COVID-19 no fluxo de atendimento, e o engajamento de pacientes e serviços na prevenção de recorrências.

Estudos recentes nas diversas áreas da computação têm demonstrado que os melhores resultados são obtidos pela combinação de diferentes técnicas, o que não é diferente no caso da mineração de processos e da saúde em especial. São exemplos deste modelo de associação um framework

integrando mineração de processos, análise multicritério e simulação [1]; e análise de processos e dados.

A despeito das pesquisas e soluções apresentadas anteriormente, é possível observar que a mineração de processos é uma técnica que tem potencial de apoiar a saúde. Porém, é importante lembrar que devemos primeiramente considerar o problema de

saúde a ser analisado e a partir disto modelar o processo, que depende da coleta, acesso e preparação dos dados. Estes desafios são existentes em qualquer estratégia de análise mas, especificamente na saúde, este desafio é potencializado tendo em vista a granularidade dos dados associados à diversidade de eventos representados em exames, procedimentos e demais eventos particulares à área.

Referências

1. Boareto, P.A.; Safanelli, J.; Liberato, R.; et al. A Hybrid Model to Support Decision Making in the Stroke Clinical Pathway. Simulation modelling practice and theory, v. 119, p. 102602-102617, 2022. doi:10.1016/j.simpat.2022.102602
2. Ciccarese, P., Caffi, E., Boiocchi, L., Halevy, A., Quaglini, S., Kumar, A., Stefanelli, M. The NewGuide Project: guidelines, information sharing and learning from exceptions. In Conference on Artificial Intelligence in Medicine in Europe _AIME 2003.. 2003. p. 163-167. Berlin, Heidelberg. Springer.
3. MantovanI, L. K.; Cintho, L. M. M.; Santos, E. A. P; Moro, C. M. C. Mineração de Processos em Atendimentos de Acidente Vascular Cerebral. In: XVI Congresso Brasileiro de Informática em Saúde - CBIS 2018, 2018, Fortaleza. Journal of Health Informatics, 2018. v. 1. p. 727-741.
4. Rosa, C. O.C.S; Ito, M.,Vieira, A. B.; Gomes, A. T. A. Modelagem, Mineração e Análise de Jornadas / Trajetórias de Pacientes. In: Lucas Ferrari de Oliveira; Flavio Henrique Duarte de Araujo. (Org.). Minicursos do SBCAS 2022. 1 ed. Porto Alegre: Sociedade Brasileira de Computação, 2022, v. 1, p. 96-145. doi: 10.5753/sbc.10508.8
5. Xavier E.C., Sutile, V.M.; Miranda C.M., Carvalho D.R. Análise de itinerário dos usuários para atendimento de serviços Básicos de Saúde, por meio de Mineração de Processos. J. Health Inform. 2021 outubro-dezembro; 13(4): 120-127



MÁRCIA ITO é orientadora no Mestrado em Sistemas Produtivos do CEETEPS. Médica (UNIFESP), tecnóloga em Processamento de Dados (FATEC-SP), doutora em engenharia elétrica (EPUSP). É Secretária Regional São Paulo Leste e membro do comitê gestor da Comissão Especial de Computação Aplicada à Saúde da SBC. Foi coordenadora geral da gestão da informação estratégica do Ministério da Saúde (2019-2020) e pesquisadora da IBM Research Brasil (2012-2018).

DEBORAH RIBEIRO CARVALHO é Professora da Pontifícia Universidade Católica do Paraná-PUCPR, com mestrado (1999) e doutorado (2002) em Informática pela PUCPR e doutorado em Computação de Alto Desempenho pela UFRJ (2005). Atua como pesquisadora em inteligência artificial aplicada à saúde nas áreas clínica e de gestão.

CLAUDIA MORO é professora titular da Pontifícia Universidade Católica do Paraná (PUCPR) no Programa de Pós-Graduação em Tecnologia em Saúde. Doutora em Engenharia Elétrica pela Universidade de São Paulo. Atua em saúde digital e inteligência artificial em saúde. Iniciou pesquisas em mineração de processos no Laboratório de Informática Biomédica "Mario Stefanelli" da Universidade de Pavia, Itália.

POSCOMP 2023

EXAME NACIONAL PARA
INGRESSO NA PÓS-GRADUAÇÃO
EM COMPUTAÇÃO

Inscrições:

24 de abril à 16 de agosto, até às 17h.

Prova: 17 de setembro.

**PARA MAIS INFORMAÇÕES
ACESSE [BIT.LY/POSCOMP2023](https://bit.ly/poscomp2023)**



CONGRESSO
DA SOCIEDADE
BRASILEIRA DA
COMPUTAÇÃO

CSBC 2023

06 A 11 DE AGOSTO

JOÃO PESSOA/PB

SAIBA MAIS SOBRE O EVENTO:
[CSBC.SBC.ORG.BR](https://csbc.sbc.org.br)

**ACESSE
[BIT.LY/CSBC_2023](https://bit.ly/csbc_2023)**





ARTIGO

MINERAÇÃO DE PROCESSOS APLICADA A PRODUÇÃO E LOGÍSTICA: DESAFIOS PARA A ADOÇÃO DA TÉCNICA PELA INDÚSTRIA

POR

*Eduardo Alves Portela Santos, Silvana Pereira Detro, Alexandre Choueri Checoli
portela@ufpr.br, silvana.detro@ufpr.br e alexandrechecoli@ufpr.br*

As operações de produção e logística correspondem a um dos domínios mais interessantes para o gerenciamento de processos de negócios. Na área de produção, observa-se a necessidade cada vez maior de diminuição dos tempos de lançamento dos produtos, a otimização de recursos de manufatura e fabricação, diminuição de custos e aumento da qualidade dos produtos. Por sua vez, a logística lida atualmente com requisitos de desempenho extremamente rígidos, com a demanda por entregas ultrarrápidas e pressão por diminuição

de custos. Esse cenário gera uma busca por melhoria de processos de forma contínua nunca antes vista.

Os desenvolvimentos de hardware e software nos últimos anos aumentaram a capacidade de gerar e armazenar informações, levando a um volume massivo de dados brutos no chão de fábrica. Essa disponibilidade de dados tem despertado interesse no desenvolvimento de novas técnicas capazes de extrair deles informações cada vez mais úteis. Nos ambientes de manufatura e logística, isso é refor-

çado pela busca da Indústria 4.0 e pelo conceito *smart supply chain*, em que as máquinas trabalham integradas como uma rede colaborativa e a cadeia de suprimentos é integrada de ponta a ponta em todas as suas operações, fornecendo aos gestores a possibilidade de usar dados on-line para tomar decisões mais precisas [3].

Apesar de sua origem nos processos de negócios, a mineração de processos tem liderado um grande movimento em direção à descoberta de modelos em ambientes de manufatura e logística nos últimos anos. No contexto industrial, a mineração de processos tem sido aplicada como método de avaliação de desempenho, orientando os gestores no ciclo de melhoria da qualidade, na gestão da manutenção, no uso integrado com a metodologia six sigma, dentre outras aplicações. Na logística, as aplicações vão desde a logística portuária até o gerenciamento da cadeia de suprimentos. A Figura 1 apresenta uma visão geral de aplicações em relação à área de operações e em relação ao setor industrial [2,1].

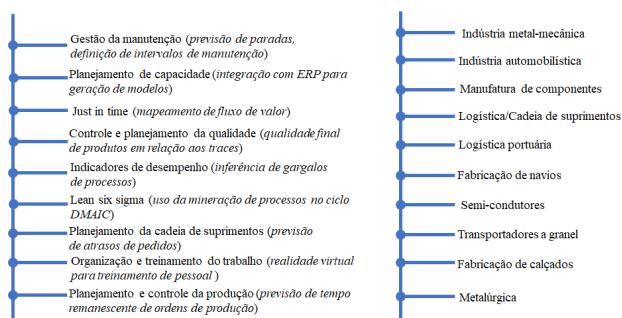


FIG. 01 | APLICAÇÕES DA MINERAÇÃO DE PROCESSOS.

Por ser uma ferramenta tão versátil, a mineração de processos pode ser utili-

zada nas mais diversas formas, uma vez que sua comunidade de pesquisa está sempre adicionando novos recursos às ferramentas de código aberto como, por exemplo, o Process Mining for Python - PM4PY¹.

Desafios

A complexidade dos processos produtivos aumenta constantemente para acompanhar os avanços da tecnologia, tendências de mercado, exigências legislativas e, acima de tudo, produtos de alta qualidade. O aumento da complexidade pode ser devido ao aumento da interação entre diferentes subsistemas, adição de funcionalidades para lidar com variantes de produtos pela estação ou máquina, adição de lógica para garantir a segurança e técnicas de otimização empregadas para melhorar fatores como tempo de processo, consumo de energia, requisitos de qualidade, etc. Enquanto novas técnicas e processos estão sendo pesquisados sob a égide da Indústria 4.0, não é fácil implementar essas técnicas em sistemas existentes ou sistemas legados [5].

Além disso, uma vez que uma linha de produção é posta em operação, espera-se que ela dure vários anos para compensar o investimento e começar a gerar retorno financeiro. Na maioria das vezes, o obstáculo de custos mais altos dificulta a adoção de novas tecnologias, portanto, a lacuna entre as tecnologias existentes no chão de fábrica e as tecnologias recentes disponíveis raramente diminui. Como consequência, o chão de fábrica consiste em uma mistura diversificada

¹ <https://pm4py.fit.fraunhofer.de/>

de tecnologias que vão desde sistemas legados até sistemas de última geração, tornando a manutenção da compatibilidade entre sistemas tecnológicos tão diversos um grande desafio. Especificamente para a aplicação da técnica de mineração de processos, essa mistura diversificada de tecnologias gera uma série de dificuldades: (i) extração do log de eventos em um ambiente heterogêneo, com múltiplos sistemas de informação, impossibilitando a uniformização dos dados; (ii) dificuldade na caracterização do processo 'ponta a ponta', uma vez que parte(s) do processo pode(m) estar sendo suportados por sistemas legados, que por sua vez podem não gerar logs de eventos ou gerar logs inadequados; e (iii) uso limitado dos algoritmos e técnicas da mineração de processos em função da limitação ou mesmo da inexistência de logs em partes do processo.

Uma outra característica do ambiente industrial impõe um outro desafio para a aplicação da mineração de processos: a definição de camadas ou níveis é uma forma bastante comum de organização deste segmento. Modelos clássicos com a 'pirâmide da automação' e documentos mais formais como a norma técnica internacional ISA 95 [4] adotam essa visão. A ISA 95, por exemplo, considera que um ambiente industrial é composto de quatro níveis, sendo o nível 0 correspondente ao sistema de produção real; o nível 1, formado pelos sensores e atuadores do processo (máquinas, equipamentos); o nível 2, composto pelos sistemas de controle e supervisão (hardware e software); o nível 3, caracterizado pelos sistemas de gerenciamento da operação de manufatura (p. ex.,

Manufacturing Execution System – MES); e o nível 4 formado pelos sistemas de planejamento empresarial e de logística (p. ex., Enterprise Resource Planning – ERP).

Esses diferentes níveis trabalham com diferentes horizontes de tempo (p. ex. o nível 4 pode trabalhar com planejamento de produção de meses enquanto os níveis 1 ou 2 podem trabalhar com horizontes de milissegundos). Na perspectiva de processos e da mineração de processos, essa característica impõe diferentes níveis de granularidade. Existem múltiplos processos rodando em diferentes horizontes de tempo. Dessa forma, podemos ter logs de eventos oriundos diretamente de uma máquina de produção, como um centro de usinagem, em que os eventos correspondem a ocorrências no equipamento tais como início e fim de operação, falha, nível de temperatura, quebra de broca, etc. Podemos também ter logs de eventos que consideram dados de tempo de produção de uma determinada célula ou planta, com múltiplas máquinas e equipamentos de produção. O log nesse caso terá dados mais abstratos, como as operações de manufatura (tornear, furar, fresar, entre outros). Podemos seguir para níveis mais acima, em que os logs de eventos podem caracterizar o processo de atendimento de ordens de produção que são realizadas em diversas células de trabalho ou o processo logístico mais amplo, que engloba a produção, estocagem, movimentação e a entrega dos produtos ao cliente.

Outro desafio relevante para a adoção da mineração de processos na indústria é a presença de diferentes sistemas de informação nas diferentes áreas. No caso,

mesmo considerando uma determinada máquina ou centro de trabalho de manufatura, é possível que as diferentes áreas da empresa lidem com conjuntos específicos de dados em diferentes sistemas de informação. Por exemplo, os gestores da manutenção geralmente lidam com eventos oriundos da máquina, chegando até o nível de componente (p. ex. nível de temperatura da ferramenta de corte, corrente no motor, etc.). Já os gestores da produção estão interessados em dados mais relacionados ao estado da máquina que caracterizam sua disponibilidade e desempenho para operar (p. ex. início e fim de operação, desligamento por falta de matéria prima, setup, parada para limpeza, entre outros).

Então, para atender às necessidades apresentadas, cada área tem os seus próprios recursos de informação. Esse aspecto acarreta uma dificuldade na observação e na análise do processo como um todo. Por exemplo, caso se queira analisar uma determinada etapa de um processo de manufatura, englobando simultaneamente eventos da máquina e de produção, certamente haverá muitas dificuldades para tal iniciativa em função das múltiplas fontes de dados. Apesar dos imensos desafios e obstáculos descritos, a área de operações e logísticas pode se beneficiar de forma extraordinária a partir do uso da mineração de processos.

Referências:

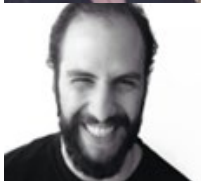
1. CORALLO, A.; LAZOI, M.; STRIANI, F. Process mining and industrial applications: A systematic literature review. *Knowledge and Process Management*, 27, 3, p. 225-233, 2020.
2. DOS SANTOS GARCIA, C.; MEINCHEIN, A.; FARIA JR., E. R.; DALAGASSA, M. R.; SATO, D. M. V.; CARVALHO, D. R.; SANTOS, E. A. P.; SCALABRIN, E. E. Process mining techniques and applications – A systematic mapping study. *Expert Systems with Applications*, v. 133, p. 260-295, 2019.
3. GARAY-RONDERO, C.L., MARTINEZ-FLORES, J.L., SMITH, N.R., MORALES, S.O.C., Aldrette-Malacara, A., 2020. Digital supply chain model in industry 4.0. *Journal of Manufacturing Technology Management*. v. 31, n. 5, p. 887-933, 2020.
4. INTERNATIONAL SOCIETY OF AUTOMATION. ANSI/ISA-95.00.03-2005 Enterprise-Control System Integration. Part 3: Activity Models of Manufacturing Operations Management, 2005.
5. RINDERLE-MA, S.; MANGLER, J. Process Automation and Process Mining in Manufacturing. In: Polyvyanyy, A., Wynn, M.T., Van Looy, A., Reichert, M. (eds) *Business Process Management, BPM 2021*, Lecture Notes in Computer Science, vol 12875. Springer, Cham, 2021.



EDUARDO ALVES PORTELA SANTOS é Professor Adjunto na Universidade Federal do Paraná. Doutor em Engenharia Elétrica pela Universidade Federal de Santa Catarina. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Eindhoven University of Technology, nos Países Baixos.



SILVANA PEREIRA DETRO é Professora Adjunta na Universidade Federal do Paraná. Doutora em Engenharia de Produção pela Pontifícia Universidade Católica do Paraná e Doutora em Automatique, Traitement du Signal et des Images, Génie Informatique pela Université de Lorraine (França). Atua em gestão de processos de negócio e mineração de processos.



ALEXANDRE CHOUZEIRI CHECOLI é Professor Adjunto na Universidade Federal do Paraná. Doutor em Engenharia de Produção pela Pontifícia Universidade Católica do Paraná. Atua em mineração de processos, otimização combinatório, heurísticas e meta-heurísticas.



ARTIGO

MINERAÇÃO DE PROCESSOS: OPORTUNIDADES PARA A TRANSPARÊNCIA PÚBLICA

POR

Renata Araujo e Marcelo Fantinato

renata.araujo@mackenzie.br e m.fantinato@usp.br

Durante a pandemia, ouvimos muito falar sobre a Anvisa – Agência Nacional de Vigilância Sanitária ¹ –, autarquia ligada ao governo federal responsável por promover a proteção à saúde por meio do controle da produção, consumo, importação e exportação de produtos, assim como do controle da qualidade de serviços de saúde no Brasil. Muitos cidadãos brasileiros reconhecem a sua importância institucional e, em linhas muito gerais, a que se destina. Mas, se perguntássemos a você, leitor, provavelmente um cidadão brasileiro, se você conhece como a Anvisa funciona, como são seus procedimentos,

¹ <https://www.gov.br/anvisa/pt-br>

o que você responderia?

Há mais de dez anos, as instituições brasileiras de administração pública realizam um esforço importante na gestão e disponibilização de informação sobre o funcionamento dos atores públicos [1]. O valor dessas iniciativas é demonstrado pela continuidade em investimento na digitalização de serviços de governo, mesmo em ambientes políticos eventualmente refratários à abertura da informação [2]. Um principal exemplo desse esforço é o Portal de Serviços e Informações do Brasil ², onde cidadãos (com acesso à internet) podem navegar e conhecer toda a sorte de serviços ofe-

² <https://www.gov.br/pt-br>

recidos pelos diversos atores governamentais. Outro exemplo é o Portal Brasileiro de Dados Abertos³, no qual, por meio de estratégias unificadas, as diversas áreas governamentais disponibilizam dados relacionados ao seu funcionamento em formato aberto. Os serviços da Anvisa e uma série de informações sobre seu funcionamento também estão lá, disponíveis aos cidadãos para acesso e uso, como resposta às obrigações da transparência pública.

A transparência é um dos aspectos essenciais do conceito de governo aberto [3, 7], e sua promoção é fundamental para o exercício da democracia ao estimular melhorias na gestão por parte das instituições públicas, assim como pela possibilidade de acompanhamento e participação ativa dos cidadãos nos resultados da gestão pública. No Brasil, a Lei de Acesso à Informação (LAI) [1] define a transparência a partir de duas dimensões: (i) transparência ativa – quando a administração pública divulga informações à sociedade por iniciativa própria, independentemente de qualquer solicitação – ; e (ii) transparência passiva – quando a administração pública divulga informações a pedido de cidadãos e da sociedade.

Você, como cidadão brasileiro, que tanto ouve falar sobre a Anvisa nos canais de mídia, conhece a abrangência de suas finalidades? Você conhece os processos que essa autarquia realiza cotidianamente? Sabia que estão disponíveis para sua análise diversos conjuntos de dados sobre processos e assuntos relacionados à vigilância sanitária no Brasil, como os ingredientes ativos de agrotóxicos no país, dados sobre

a venda de medicamentos controlados, sobre a produção em centros de reprodução humana assistida e sobre as petições de registros de produtos de higiene, perfumes, cosméticos, saneantes, entre tantos outros temas ligados ao nosso dia a dia?

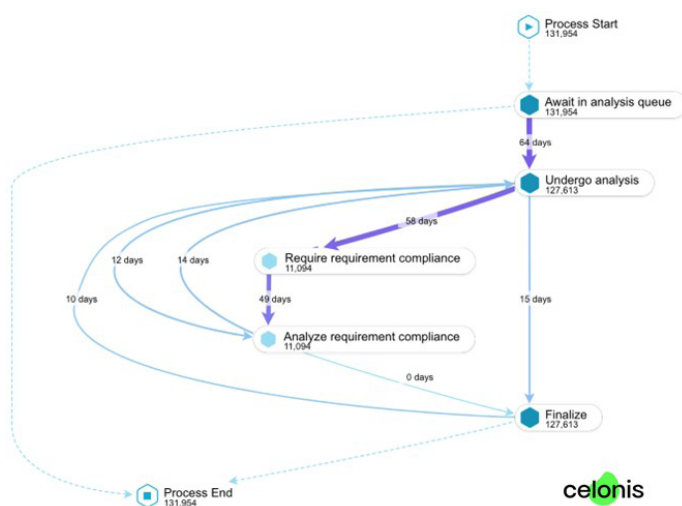
Embora tenhamos a LAI como um importante instrumento regulatório em nosso país, que permite que as informações como as descritas acima sejam disponibilizadas ao cidadão, as informações para a prática da transparência pública são muito mais utilizadas pelas agências de controle da gestão pública e estão ainda longe de serem naturalmente utilizadas e incorporadas tanto na cultura interna das organizações como no cotidiano dos cidadãos. A prática da transparência precisa ser continuamente incentivada e aperfeiçoada, considerando principalmente as possibilidades emergentes de novas tecnologias, incluindo as relacionadas à gestão da informação [5].

Nesse intento, vemos uma boa oportunidade para abordar e explorar a mineração de processos como prática para a transparência de processos no âmbito da administração pública, incentivando respostas inovadoras no sentido da gestão dos processos e da informação para transparência; como forma de explorar tecnologias avançadas para o exercício da transparência; como auxílio no entendimento por parte dos servidores públicos do funcionamento de suas atividades; como um meio em potencial para mudança de cultura interna; e como um canal em potencial para a comunicação com os cidadãos sobre os processos executados no atendimento aos serviços públicos.

³ <https://dados.gov.br/home>

Como exemplo das oportunidades de mineração de processos para a transparência de processos de prestação de serviços públicos, exploramos [6] a análise dos dados do processo “Petições para Registro de Produtos de Saúde” da Anvisa, registrados no log de eventos do DataVisa (sistema de informação que apoia os processos da agência), disponibilizado no Portal de Dados Abertos. A análise descritiva do processo permitiu levantar indicadores de execução, detalhes sobre o volume de trabalho e pontos de atenção sobre a eficiência do processo. A descoberta do modelo de processo revelou informações sobre o fluxo real de execução do processo (Figura

1), incluindo as variantes mais comuns, casos potencialmente abertos por permanecerem em análise indefinidamente, e o gargalo na análise da petição quando há requisições e análise de cumprimento de exigências. O modelo de processo descoberto automaticamente apresentou 80% de conformidade com o modelo de processo descrito em normativa da Anvisa, com a seguinte violação mais comum identificada: embora “Análise de cumprimento de exigência” devesse ocorrer após “Requisição de cumprimento de exigência”, verificou-se que há casos em que a análise é realizada sem uma requisição prévia.



Legenda:

- Process Start: Início do processo
- Await in analysis queue: Espera em fila de análise
- Undergo analysis: Análise em andamento
- Require requirement compliance: Requisição de cumprimento de exigência
- Analyze requirement compliance: Análise de cumprimento de exigência
- Finalize: Finalização
- Process End: Término do processo

Grafo de fluxo direto obtido filtrando os cinco *traces* (ou variantes de processo) mais comuns, anotado com a frequência em que as atividades são executadas e o tempo médio transcorrido entre as execuções das atividades.

Estatísticas descritivas e observações básicas sobre o processo completo (todas as variantes):

- 135 eventos ocorrem por dia (em média)
- 30 novos casos iniciados por dia (em média)
- pico de novos casos (111 casos) observado em maio de 2010
- 777 variantes observadas
- variante mais comum (“espera em fila de análise” – “análise em andamento” – “finalização”) observada em 47,64% dos casos, com duração média de 49 dias
- intercorrência mais comum (“requisição e análise de cumprimento de exigência”) em 17% dos casos
- gargalo observado entre “espera em fila de análise” e “análise em andamento” e entre “análise em andamento” e “requisição e análise de cumprimento de exigência”)



FIG. 01 | RESULTADO DA DESCOBERTA DE MODELO DE PROCESSO DA ANVISA [6]

Partimos do pressuposto de que só é possível participar de decisões públicas e da inovação do serviço público quando os cidadãos conhecem como a administração pública funciona [4], o que pode ser oferecido por modelos descritivos criados pela mineração de processos. Do contrário, as percepções dos cidadãos sobre a administração pública ficarão restritas ao que funciona ou não, em sua perspectiva

de beneficiário, sem considerar como, porque e quais as razões para esses resultados. Além disso, a mineração de processos possibilita à administração pública estabelecer ações e políticas eficazes de melhoria de procedimentos com base em modelos preditivos e prescritivos orientados a processos. A Figura 2 resume as oportunidades de mineração de processos para transparência pública.

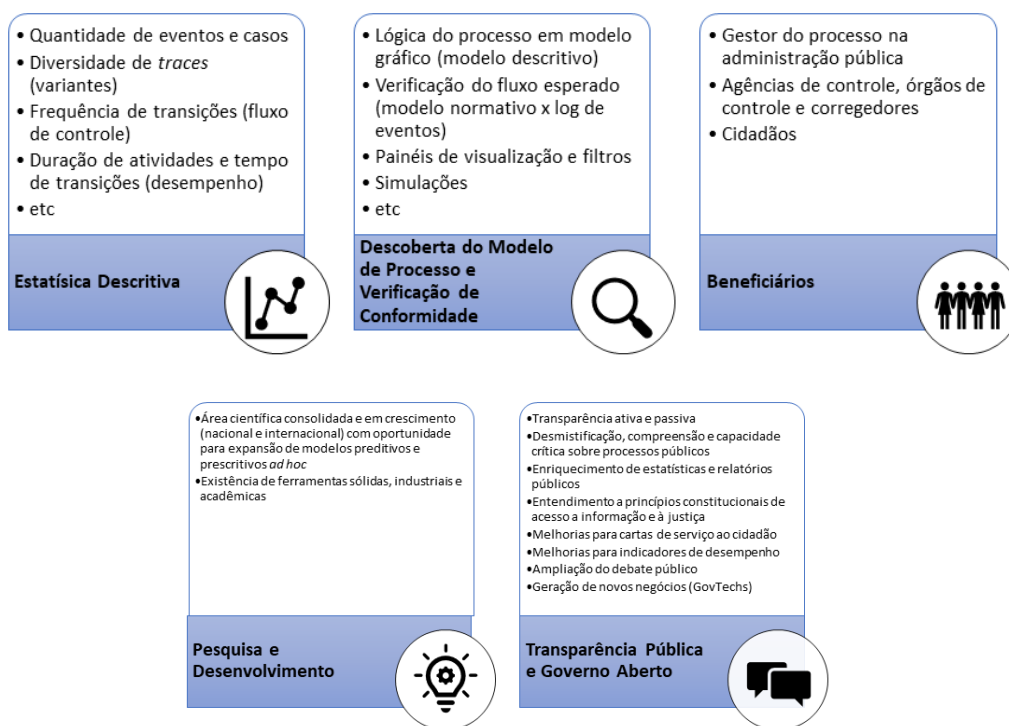


 FIG. 02 | OPORTUNIDADES DE MINERAÇÃO DE PROCESSOS PARA TRANSPARÊNCIA PÚBLICA

Claro que há também muitos desafios para aplicar mineração de processos na transparência pública: a compreensão da transparência como um valor organizacional e um objetivo de gestão a ser considerado em todo o processo de geração e uso de informação dentro e fora das organizações públicas; a dependência

da qualidade dos dados sobre os processos registrados na organização; o esforço na integração de informações entre diferentes atores públicos; a necessidade de capacitação técnica para uso das ferramentas de mineração de processos pelos profissionais da administração pública; o investimento em aquisição de ferramen-

tas; e a fundamental e necessária adaptação e simplificação das informações técnicas obtidas por meio da mineração de processos para descrições adequadas à compreensão e consumo pelo cidadão brasileiro.

Referências

1. BRASIL. LEI n. 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal. Brasília, DF: Diário Oficial da União, 2011.
2. BRASIL. LEI n. 14.129, de 29 de março de 2021. Dispõe sobre princípios, regras e instrumentos para o Governo Digital e para o aumento da eficiência pública. Brasília, DF: Diário Oficial da União, 2021.
3. LAHTROP, D.; RUMA, L. Open government: collaboration, transparency, and participation in practice, 1st ed., O'Reilly Media, 2010.
4. LAITINEN, I.; KINDER, T.; STENVALL, J. Co-design and action learning in local public services. J. of Adult and Continuing Educ., v. 24, n. 1, p. 58-80, 2018.
5. MOTA FILHO, H. E. C. A governança pública da informação: transparência e segurança jurídica. Cadernos Jurídicos, v. 2, n. 3, p. 28-40, 2022.
6. NEUBAUER, T. R.; ARAUJO, R. M.; PERES, S. M., FANTINATO, M. Transparency promoted by process mining: an exploratory study in a public health product management process. Workshop de Computação Aplicada a Governo Eletrônico, 10ª ed., p. 37-48, 2022.
7. OGP. Open Government Partnership, 2023. Disponível em: <https://www.opengovpartnership.org>. Acesso em: 07/04/2023.



RENATA ARAUJO é Professora na Faculdade de Computação e Informática e no Programa de Pós-Graduação em Computação Aplicada da Universidade Presbiteriana Mackenzie. É também pesquisadora no Programa de Pós-Graduação em Sistemas de Informação da EACH/USP e no Doutorado em Políticas Públicas da Escola Nacional de Administração Pública. Atua na área de Sistemas de Informação, Governança Digital e Ciberdemocracia. Coordena o Grupo de Pesquisa e Inovação em Ciberdemocracia, com foco em pensar novas formas de diálogo e colaboração entre pessoas e instituições com o uso da tecnologia digital.



MARCELO FANTINATO é Professor Associado na Escola de Artes, Ciências e Humanidades da Universidade de São Paulo. Doutor em Ciência da Computação pela Universidade Estadual de Campinas e Livre-docente em Gestão de Processos de Negócio pela Universidade de São Paulo. Atua em gestão de processos de negócio e mineração de processos. Foi pesquisador visitante, trabalhando com mineração de processos, na Vrije Universiteit Amsterdam e na Utrecht University, nos Países Baixos.



ARTIGO

MINERAÇÃO DE PROCESSOS NO JUDICIÁRIO BRASILEIRO

POR

Ricardo Massa F. Lima, Raphael J. D'Castro, Adriano L. I. Oliveira, Rafael L. Paulo, Bráulio G. Gusmão e João Thiago de F. Guerra
rmfl@cin.ufpe.br, raphael.dcastro@tjpe.jus.br, alio@cin.ufpe.br,
rafael.leite@trfl.jus.br, brauliogusmao@trt9.jus.br e joao.guerra@cnj.jus.br

Imagine uma instituição com presença em todos os estados da federação brasileira, que recebe mais de 25 milhões de demandas a cada ano, cada uma das quais possuindo peculiaridades próprias, que precisam ser analisadas por um ser humano, interpretadas precisamente por essa pessoa e exaustivamente debatidas com os interessados. O resultado é a produção de uma decisão fundamentada na legislação que afetará a vida de seres humanos e/ou organizações. Essa instituição existe. Trata-se do Poder Judiciário brasileiro, composto por 92 órgãos, elencados no art. 92 da Constituição da República Federativa do Brasil de 1988.

Diante de tamanha complexidade, é fundamental dispor de instrumentos para

análise e melhoria do judiciário. Um exemplo, alinhado com o conceito de Jurimetria [3], é o Relatório Justiça em Números [1], publicado anualmente pelo Departamento de Pesquisas Judiciárias (DPJ) do Conselho Nacional de Justiça (CNJ). O relatório tem foco em aspectos relacionados ao eixo produtividade, gerando indicadores como o Índice de Produtividade Comparada da Justiça (IPC-Jus).

Além dessa iniciativa, desenvolver novas tecnologias é fundamental para dar suporte às demandas do Judiciário brasileiro. Um exemplo disso é o emprego da mineração de processos para diagnóstico do funcionamento e identificação de oportunidades de melhorias no fluxo processual, bem como mensuração dos efei-

tos das mudanças. Estas medidas são capazes de contribuir para o aumento da produtividade e redução dos custos operacionais através da eliminação de etapas improdutivas, identificação e tratamento de desvios nos fluxos processuais e qualificação da atuação orientativa das inspeções judiciais.

Para avaliar os benefícios que a mineração de processos pode trazer para o Poder Judiciário, Unger et al. [4] apresentaram uma prova de conceito utilizando o log de eventos do Tribunal de Justiça do Estado de São Paulo. O trabalho utilizou a ferramenta Everflow¹ para identificar as atividades mais frequentes e os gargalos nos fluxos dos processos judiciais, demonstrando o potencial da mineração de processos para encontrar as causas de eventuais ineficiências.

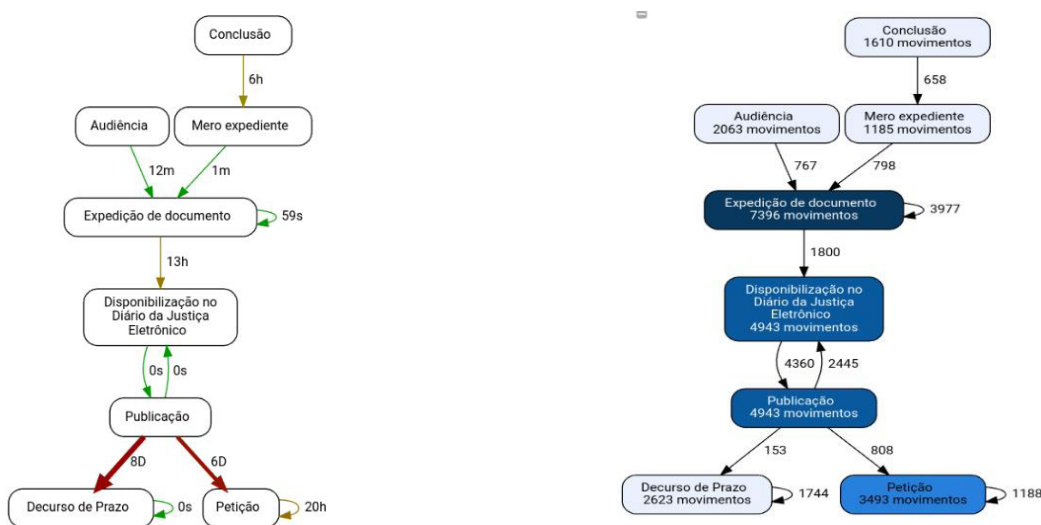
Autoridades do Poder Judiciário, como o Juiz Bráulio Gusmão, reconhecem que a mineração de processos oferece oportu-

nidade para aprimorar a forma de tratamento dos seus dados digitais por meio da aplicação de algoritmos para descoberta de fluxos de processos judiciais, conformidade e análise preditiva [2].

Inspirado em tais perspectivas, em 2020, juntamente com pesquisadores do Centro de Informática da UFPE (CIn-UFPE), Gusmão conduziu estudos no CNJ para aplicação da mineração de processos aos dados de processos judiciais armazenados no repositório de dados Codex. Como resultado, ao final daquele ano, teve início o desenvolvimento da ferramenta JuMP (Judiciário com Mineração de Processos), que explora os dados do Codex² para gerar análises que contemplam a dimensão de processo ponta a ponta com o objetivo de apresentar evidências para a tomada de decisões táticas e operacionais.

2 <https://www.cnj.jus.br/sistemas/plataforma-codex/>

1 <http://everflow.ai>



(a) Fluxo Processual com ênfase no desempenho

(b) Fluxo Processual com ênfase na frequência

FIG. 01 | FLUXO PROCESSUAL DESCOBERTO PELO JUMP, COM DESTAQUE PARA O DESEMPENHO E FREQUÊNCIA

O JuMP permite escolher, na base de dados do Codex, a unidade judicial a ser analisada e aplicar filtros para a escolha do intervalo de tempo, processos judiciais com um determinado número de movimentações, classe e assunto processual, entre outros. O sistema aplica técnicas de descoberta de processos para apresentar o modelo com o fluxo de processos de negócios presente no log de eventos, dando destaque à dimensão temporal - Figura 1(a) -, revelando os gargalos nos processos judiciais, ou à dimensão de frequência - Figura 1(b) -, com ênfase nos movimentos mais comumente realizados. Inicialmente, o JuMP apresenta uma versão simplificada do modelo de fluxo de processos. Porém, é possível aumentar a complexidade do fluxo a ser visualizado. Também são apresentados dados estatísticos sobre o log de eventos selecionados. Dentre as estatísticas exibidas estão a quantidade de processos judiciais na base, número de variantes de fluxos de processos, mediana da dura-

ção dos processos judiciais e de movimentos por processo judicial.

Na aba Processos, é possível listar e buscar processos judiciais de interesse, apresentando informações sobre a quantidade de ocorrências de cada movimentação realizada no processo judicial e a sequência de movimentos em ordem cronológica, dando destaque aos pontos de maior custo temporal (vide Figura 2). É possível ainda ordenar processos da unidade judicial escolhida em ordem crescente ou decrescente do número de movimentos realizados pelos processos judiciais ou pela duração total dos processos. Esta aba também exibe as variantes (grupo de processos judiciais que seguem o mesmo fluxo) encontradas no log. Também pode-se visualizar a quantidade de vezes que cada tipo de movimento aconteceu no processo judicial.

A aba de movimentos apresenta diversas perspectivas sobre as movimentações presentes na base de dados, incluindo gráficos de duração e de frequência, por exemplo. Já a aba Pessoas traz uma série de inova-

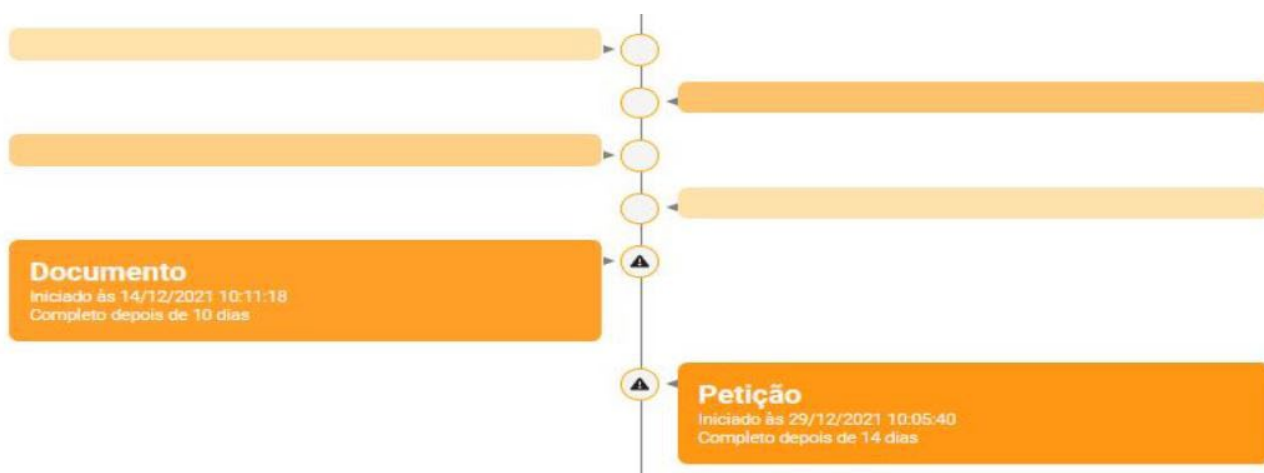


FIG. 02 | LINHA DO TEMPO DOS MOVIMENTOS REALIZADOS POR UM PROCESSO ESPECÍFICO

ções associadas à mineração de processos. Ela tem como objetivo obter informações sobre as contribuições e forma de trabalho de cada integrante da unidade judicial escolhida. Além de visualizar o fluxo de trabalho de um membro da unidade, com destaque para os movimentos realizados com mais frequência ou para o tempo gasto em cada tipo de atividade, é possível comparar a contribuição de cada pessoa em relação à média de contribuição das pessoas da unidade ao longo do tempo (vide Figura 3). Esta aba oferece ainda a oportunidade de explorar a rede social de relacionamentos funcionais de cada indivíduo, com ênfase para o contexto em que cada atividade foi realizada (atividade anterior e posterior) e para as pessoas envolvidas naquele contexto, além do tempo despendido em cada contexto e ambiente de relacionamento funcional.

Finalmente, para conduzir uma análise

de conformidade define-se um conjunto de marcos processuais em uma determinada ordem (ou seleciona-se algum template predefinido) e o JuMP cria o modelo do fluxo de processo especificado, dando destaque para a quantidade de processos judiciais que seguiram o fluxo esperado e para aqueles que apresentaram fluxo distinto. Nesse último caso, é mostrado o fluxo alternativo seguido. A visualização dá ênfase ao tempo médio consumido em cada etapa.

Desde dezembro de 2022, o JuMP está disponível para uso por integrantes dos tribunais brasileiros através da Plataforma Digital do Poder Judiciário Brasileiro (PDPJ-Br³). Seu desenvolvimento contou com a participação ativa do Juiz Auxiliar da Presidência do CNJ Rafael Leite Paulo. Desde então, o JuMP foi acessado por mais de 2.700 usuários diferen-

3 Plataforma Digital do Poder Judiciário Brasileiro – PDPJ-Br - Portal CNJ: shorturl.at/CEJUZ

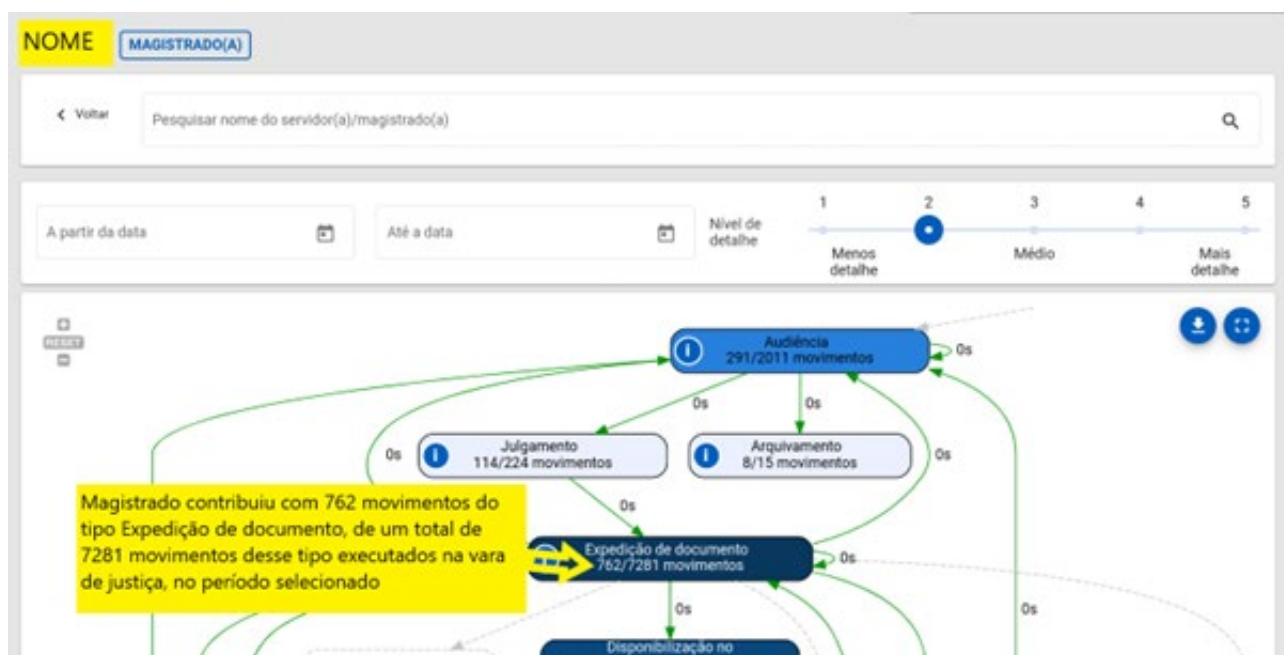


FIG. 03 | CONTRIBUIÇÃO DO MAGISTRADO NOME EM RELAÇÃO À MÉDIA DE CONTRIBUIÇÃO DOS DEMAIS MEMBROS DA UNIDADE AO LONGO DO TEMPO

tes. Os relatos dos que têm utilizado a ferramenta são muito positivos.

Atualmente, o projeto JuMP está sob a supervisão do Juiz Auxiliar da Presidência do CNJ João Thiago Guerra. Sua contribuição tem fomentado o desenvolvimento de visualizações de processos de negócio intuitivas, contextualizadas para o Judiciário, que requerem pouco esforço para transmitir a essência dos logs de eventos, sem exigir profundo conheci-

mento sobre ciência de dados, ciência de processos e/ou mineração de processos. O projeto tem sido conduzido no sentido da construção de métricas que permitam a comparação do processo de trabalho de múltiplas unidades judiciais a fim de estabelecer um fluxo processual de referência, com conseqüente identificação de fluxos desviantes, graus de desvio e pontos de otimização em relação ao modelo de referência.

Referências

1. Conselho Nacional de Justiça (CNJ). Justiça em Números 2022, shorturl.at/eliTY, acessado em 20 de janeiro de 2022.
2. Gusmão, B. Mineração de processos e gestão de casos no judiciário. Inteligência Artificial e Direito Processual: os impactos da virada tecnológica no direito processual (2 ed.), Salvador, 589–594, 2021. 4
3. Loevinger, L. Jurimetrics—The Next Step Forward. Minnesota Law Review 33, 5, 455–493, 1948. 1
4. Unger, A.J., Santos Neto, J.F., Fantinato, M., Peres, S.M., Trecenti, J., Hirota, R. Analysis of a Brazilian Court's Judicial Performance in the Business Law Processing, International Conference on Artificial Intelligence in Law (ICAIL), Copenhagen, 2022. 3



RICARDO MASSA é Professor Associado do Centro de Informática, UFPE. Atua na área de mineração de processos e avaliação de desempenho. Publicou mais de 100 artigos científicos em congressos e periódicos.



RAFAEL D'CASTRO é Secretário Adjunto de TIC do Tribunal de Justiça de Pernambuco. Atualmente faz Pós-Doutorado no Centro de Informática, UFPE, com ênfase no desenvolvimento da ferramenta JuMP.



ADRIANO L. I. OLIVEIRA é Professor Associado do Centro de Informática, UFPE. Atua na área de inteligência artificial e mineração de processos. Coordena o projeto JuMP na UFPE.



RAFAEL LEITE PAULO é Magistrado do Tribunal Regional Federal da 1ª Região e coordenou o projeto JuMP durante sua gestão como Juiz Auxiliar da Presidência do Conselho Nacional de Justiça (CNJ) entre os anos de 2000 e 2022, tendo participado ativamente da construção da ferramenta.



BRÁULIO GABRIEL GUSMÃO é juiz do Tribunal Regional do Trabalho da 9ª e Secretário-Geral do Conselho Superior da Justiça do Trabalho. Foi um dos idealizadores Laboratório de Mineração de Processos do Judiciário enquanto Juiz Auxiliar da Presidência do CNJ.



JOÃO THIAGO DE FRANÇA GUERRA é magistrado do Tribunal de Justiça do Estado de Mato Grosso (TJMT) e Juiz Auxiliar da Presidência do Conselho Nacional de Justiça na Gestão 2022/2023. Mestre em Direito e Poder Judiciário pelo Programa de Pós-Graduação Profissional em Direito da Escola Nacional de Formação e Aperfeiçoamento de Magistrados (PPGPD/Enfam). Atualmente é coordenador do projeto JuMP no CNJ.



ARTIGO

MINERAÇÃO DE PROCESSOS EDUCACIONAIS

POR

Petrônio Cândido de Lima
petronio.candido@ifnmg.edu.br

Muito antes da pandemia da COVID-19 ser declarada, em fevereiro de 2020, a Educação à Distância já era uma tendência mundial. Contudo, foi durante o período de isolamento social da COVID-19 que vários atores do processo pedagógico - professores, tutores, alunos, gestores e pais - foram subitamente forçados a adotar as tecnologias da Educação à Distância sem o devido preparo, evidenciando que as particularidades dessa modalidade de ensino podem ser desafiadoras para aqueles vindos diretamente da sala de aula tradicional.

Diversas tarefas corriqueiras do dia a dia

do professor, como o acompanhamento do comportamento do aluno, a percepção da sua evolução nas disciplinas e suas dificuldades ficam comprometidas no ambiente do ensino à distância, sendo muitas vezes relegadas a relatórios pouco intuitivos e demasiadamente abstratos. Por outro lado, ter o poder de acompanhar a forma como os alunos interagem com as atividades do curso on-line, como eles se movimentam entre os materiais didáticos disponíveis e as atividades avaliativas, pode revelar padrões de comportamento que auxiliam os professores, tutores e gestores educacionais na produção de estratégias de ensino que maximizem a aprendizagem e redu-

zam a evasão e insucesso.

Nesse contexto, a Mineração de Processos Educacionais - EPM (*Educational Process Mining*) [2] surge como uma alternativa para a extração de conhecimento em bases de dados de logs gerados pelas plataformas de ensino à distância. A EPM é uma sub-área da Mineração de Dados Educacionais [1], e pode ser realizada sob a perspectiva de múltiplos atores do processo de ensino-aprendizagem on-line: gestores e técnicos-administrativos, professores e tutores, alunos, pais e responsáveis, enfim, todos os que interagem com o sistema de ensino virtual. Cada ator representa uma perspectiva/visão do processo, um nível de acesso a diferentes ferramentas, dado que nem todos os recursos da plataforma de ensino estão acessíveis para todos os atores, logo os objetivos da mineração de processos variam de acordo com o perfil dos usuários. Para os perfis administrativos, como os professores e tutores, a conformidade de processos pode exercer um papel de destaque, mas certamente o perfil discente é o principal foco de interesse da literatura e também objeto do presente trabalho.

Os Ambientes Virtuais de Aprendizagem (*Virtual Learning Environments* - VLE) são os softwares responsáveis principalmente por auxiliar na gestão dos objetos de aprendizagem (os recursos pedagógicos) pelos professores e tutores estruturar logicamente esses objetos em unidades de ensino (cursos, módulos, disciplinas, etc) e gerenciar o acesso dos alunos, registrando e monitorando suas interações. Dentre os exem-

plos de VLEs encontram-se o software de código aberto Moodle¹ e a plataforma Google Classroom².

Um curso on-line, em sua estrutura, é uma composição de vários objetos de aprendizagem ordenados logicamente de acordo com um conteúdo programático gerenciado pelo VLE. Já os objetos de aprendizagem são os recursos didáticos do ambiente virtual que incluem páginas de texto, downloads de arquivos, atividades, provas, salas de bate-papo, fóruns, vídeos, etc., que auxiliam o professor na confecção do conteúdo do curso e em suas avaliações.

As interações dos alunos com os objetos de aprendizagem do curso geram os eventos registrados pelo VLE nos logs, produzindo os traços ou variantes do processo de negócio. Via de regra, cada aluno gera uma única variante do processo, a não ser que ele repita o curso. Cada variante desse processo deverá ser associada ao status de término, que corresponde às situações finais do aluno no curso: aprovado, reprovado, desistente, etc.

Partindo desses componentes - VLEs, cursos e logs - é possível delinear uma metodologia de EPM [3], sintetizada na Figura 1 a partir de uma simplificação do processo tradicional de descoberta de conhecimento em bases de dados. Essa metodologia constitui-se de quatro etapas: Aquisição de Dados, Pré-Processamento de Dados, Rotulação de Casos e Mineração de Processos. As etapas e suas atividades são detalhadas a seguir.

1 <https://moodle.com/pt/>

2 <http://classroom.google.com>

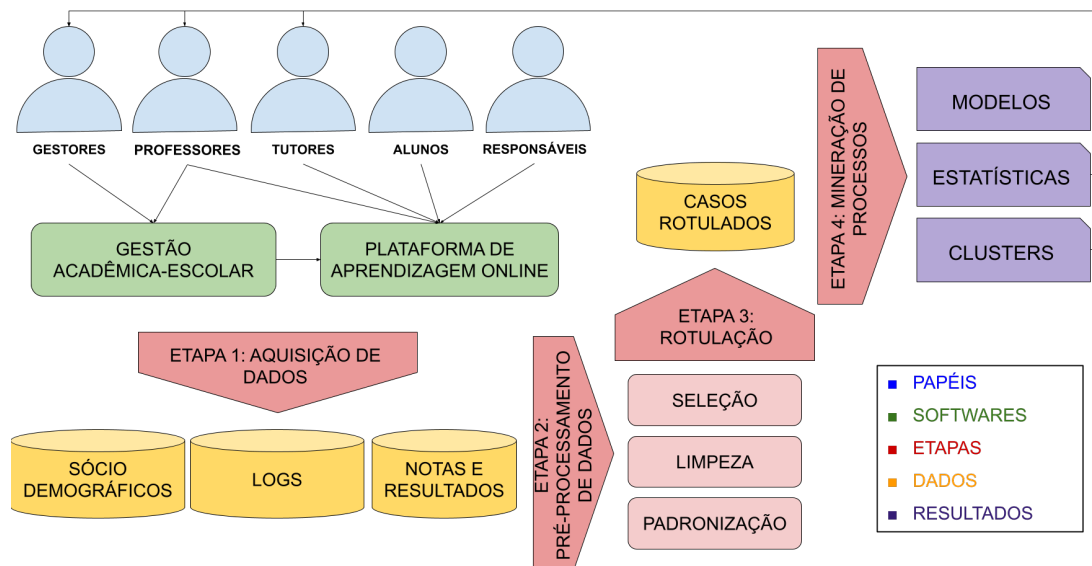


FIG. 01 | SÍNTESE DE UMA METODOLOGIA PARA MINERAÇÃO DE PROCESSOS EDUCACIONAIS

A etapa inicial, **Aquisição de Dados**, compreende a obtenção de todos os logs do VLE referentes à unidade de ensino sob estudo. A Aquisição de Dados deve ser enriquecida com dados auxiliares como as notas e resultados finais e, opcionalmente, com dados sócio-demográficos dos alunos, ampliando as possibilidades de descoberta de conhecimento. No entanto, é importante salientar que os dados podem ser classificados como informações sensíveis pela legislação vigente de proteção de dados pessoais e sua utilização indevida pode levar a sanções penais. Diante do exposto, é salutar que os dados coletados no VLE ou nos Sistemas de Gestão Acadêmica sejam pré-processados para a anonimização e remoção de informações sensíveis. Em particular, as notas e situações dos alunos nos cursos são importantes para classificar cada caso do processo, e posteriormente distinguir os comportamentos típicos nos casos que podem ser associados a cada situação final.

A etapa de **Pré-Processamento de Dados** compreende as atividades de sele-

ção, limpeza e padronização dos logs. Na etapa de seleção, os logs devem ser filtrados selecionando apenas as atividades e usuários correspondentes aos perfis desejados, nesse caso o dos alunos. Nessa etapa é importante também decidir qual a granularidade de análise dos eventos, que implicará a granularidade de representação dos objetos de aprendizagem. Isso implica em escolher se os eventos dos objetos de aprendizagem serão tratados individualmente, gerando uma atividade específica, ou se eles serão agrupados por tipo de objeto, aumentando a granularidade, simplificando o processo final, porém perdendo um pouco de poder de análise.

A etapa de **Rotulação dos Traços** é caracterizada pelo cruzamento de dados entre os casos e os dados de notas e resultados para a rotulação do processo entre as situações finais. Essa etapa é crucial para que se possa discriminar aspectos específicos nos modelos de processo que sejam característicos de uma situação final de curso. Essa etapa é igualmente mandatória quando houver a demanda

de aplicação de algoritmos supervisionados de aprendizado de máquina, particularmente visando tarefas de classificação e predição. Além das situações finais, os logs podem ser rotulados com dados sócio-demográficos dos alunos, se disponíveis, que auxiliarão na identificação de comportamentos correlacionados com cada rótulo.

Na quarta etapa, a de **Mineração de Processos**, os logs já pré-processados e rotulados são combinados e destilados para extrair um modelo de processos de negócio que representa o curso on-line e as estatísticas relacionadas a esse modelo. Diversas atividades de mineração de processos podem ser empregadas [4], dentre elas a Descoberta e Simplificação de Processos, responsável por extrair o processo a partir dos logs, Conformidade de Processos, que valida a adequação de um caso ao modelo de processo gerado, etc. As atividades dessa etapa pode ser inteiramente desempenhadas utilizando ferramentas computacionais tradicionais da Mineração de Processos, como o P²RoM³, pm4py⁴, Apromore⁵, Celonis⁶, dentre outras.

O modelo do processo, descoberto a partir dos logs, deve ser objeto de estudo dos professores e gestores educacionais. É importante contrastá-lo com a estrutura do curso no VLE, que ainda pode ser realizado a partir dos métodos de *Compliance*. Essa análise é uma ferramenta útil para detectar disparidades entre o

planejamento do professor e a real forma como os alunos consomem o curso. O Agrupamento de Casos (clustering) é empregado na identificação de perfis de comportamento dos alunos. Para investigações mais detalhadas sobre ocorrências em casos isolados, o Replay de Casos é uma ferramenta valiosa para os professores e tutores compreenderem a dinâmica e decisões do estudante.

Outro resultado do processo de mineração são as estatísticas de processo [3] [5], que incluem as distribuições de frequência de eventos (a frequência em que uma atividade do processo é acessada, considerando o processo geral ou cada caso), a distribuição de tempos gastos pelos alunos em cada atividade do processo, entre outras. Essas distribuições nos permitem identificar, em suas caudas, tanto os objetos de aprendizagem irrelevantes (que nunca ou quase nunca são acessados) quanto os gargalos, aquelas atividades que prendem o aluno por muito tempo, bloqueando a continuidade do curso. A identificação desses objetos é crucial para auxiliar o professor a otimizar o curso on-line.

Por fim, conclui-se que a metodologia da EPM leva à extração de uma forma de conhecimento valiosa no contexto educacional: o comportamento. Cada traço de um aluno é uma amostra da dinâmica de aprendizagem, que será representada em um modelo de processo, com suas estatísticas e agrupamentos. Embora seja uma metodologia valorosa, o seu emprego ainda não é difundido, existindo ainda oportunidades de pesquisa em aberto.

3 <https://promtools.org/>

4 <https://pm4py.fit.fraunhofer.de/>

5 <https://apromore.com/>

6 <https://www.celonis.com/>

Referências:

1. ALEEM, Abdul; GORE, Manoj Madhava. Educational data mining methods: A survey. In: 2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT), pp. 182-188. IEEE, 2020.
2. BOGARÍN, Alejandro; CEREZO, Rebeca; ROMERO, Cristóbal. A survey on educational process mining. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v. 8, n. 1, p. e1230, 2018.
3. DAYRELL, Marcella Souza; SILVA, Petrônio Cândido de Lima. Análise do Comportamento Discente em Cursos à Distância Baseada em Mineração de Processos. In: Anais do Encontro Nacional de Modelagem Computacional, Encontro de Ciência e Tecnologia de Materiais, Conferência Sul em Modelagem Computacional e Seminário e Workshop em Engenharia Oceânica. Anais...Pelotas(RS) UFPel / FURG / UNIPAMPA, 2022.
4. GARCIA, Cleiton dos Santos et al. Process mining techniques and applications—A systematic mapping study. Expert Systems with Applications 133 (2019): 260-295.
5. NETO, José F. S.; PERES, Sarajane M., CORREIA, Paulo; FANTINATO, Marcelo. Is my classroom flipped? using process mining to avoid subjective perception. eLearn, 2021(12).



PETRÔNIO CÂNDIDO DE LIMA E SILVA é professor no Instituto Federal do Norte de Minas Gerais. Doutor em Engenharia Elétrica pela Universidade Federal de Minas Gerais. Atua na área de Ciência de Dados e Inteligência Computacional. Atua nos grupos de pesquisa {cidic}/IFNMG, MINDS/UFMG e no PPGMCS/Unimontes.

Associe-se ou renove sua associação

Continue fortalecendo a nossa Sociedade e usufrua dos privilégios de fazer parte da maior Sociedade de Computação da América Latina!

Valores:

- Efetivo/Fundador: R\$ 336,48
- Professora e Professor da Educação Básica: R\$ 118,18
- Estudante de Pós-graduação: R\$ 118,18
- Estudante: R\$ 28,94
- Institucional: R\$ 3.235,67

**PARA SE ASSOCIAR OU RENOVAR ACESSE
CENTRALDESISTEMAS.SBC.ORG.BR/MOM**

Para mais informações, estamos disponíveis no e-mail sbc@sbc.org.br ou no telefone (51) 99252-6018 (também WhatsApp).





Sociedade Brasileira
de Computação

sbc.org.br