



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](mailto:sarajane@usp.br), [m.fantinato@usp.br](mailto:m.fantinato@usp.br) e [portela@ufpr.br](mailto:portela@ufpr.br)

**I**magine 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.

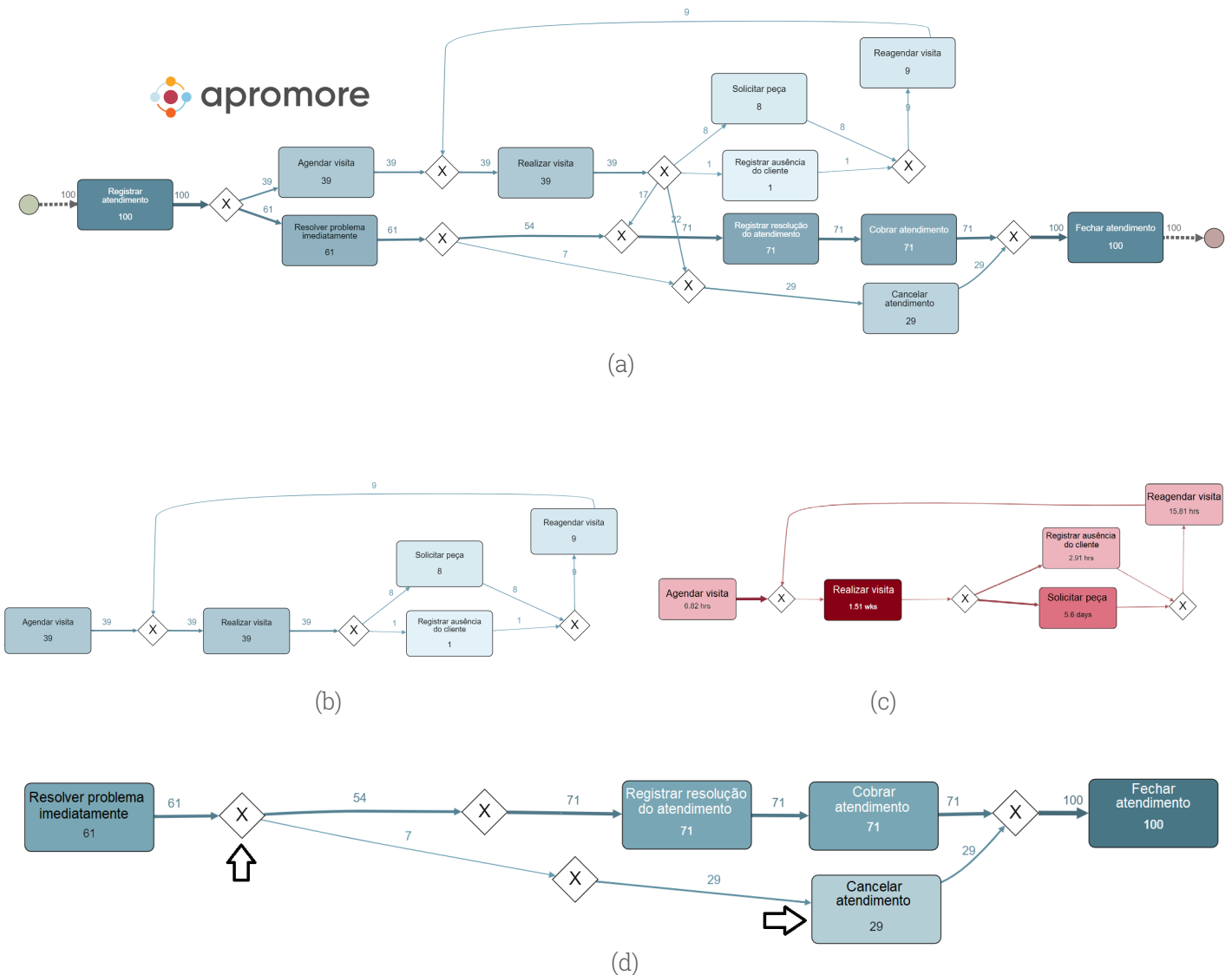
Vamos clarear essa discussão considerando o log de eventos apresentado na Figura 1. O resultado de um algoritmo de descoberta de modelo é mostrado na Figura 2a (em notação BPMN<sup>1</sup>) e nos dá a visão geral do fluxo de trabalho realizado no processo segundo o que foi registrado no log de eventos. 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.

$\sigma$

**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.

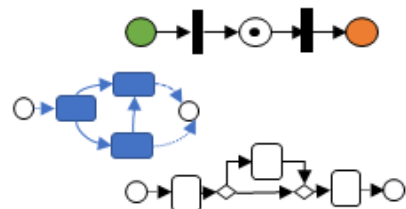


Na prática de mineração de processos, é bastante comum encontramos a denominação **variante** para uma determinada sequência de atividades existente no log de eventos. De fato, uma *trace* simples diz respeito a uma variante. Ainda, por vezes, usa-se a denominação *trace* para indicar o trace simples,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



**Notações para representação mais comuns**

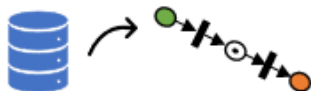
- **Petri Net** (e uma variação especial chamada *Workflow-net*).
- **DFG** – *Direct Flow Graph*.
- **BPMN** – *Business Process Model Notation*.



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

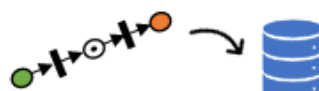
**play in**

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** ( $\alpha$ -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)

