

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.

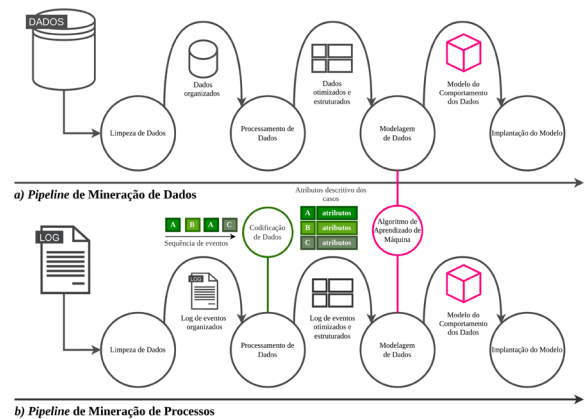


FIG. 01 | COMPARAÇÃO ENTRE OS PIPELINES PARA MINERAÇÃO DE DADOS E MINERAÇÃO DE PROCESSOS

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.