UNIVERSIDADE FEDERAL DE JUIZ DE FORA CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

T	ATIDA	MATOC	VASCONCEL.	TOC
	.AUKA	WIA IUS	VASCUNCEL	1 (1)

Metodologia para geração de registro de eventos artificial a partir da simulação de eventos discretos.

JUIZ DE FORA

LAURA MATOS VASCONCELLOS

Metodologia para geração de registro de eventos artificial a partir da simulação de
eventos discretos

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheira de Produção.

Orientador: D.Sc., Antônio Ângelo Missiaggia Picorone

JUIZ DE FORA

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pela autora

Matos Vasconcellos, Laura.

Metodologia para geração de registro de eventos artificial a partir da simulação de eventos discretos / Laura Matos Vasconcellos. -- 2020.

51 f.

Orientador: Antônio Ângelo Missiaggia Picorone Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Juiz de Fora, Faculdade de Engenharia, 2020.

1. mineração de processos. 2. registro de eventos. 3. modelo de processos. 4. simulação de eventos discretos. I. Ângelo Missiaggia Picorone, Antônio, orient. II. Título.

LAURA MATOS VASCONCELLOS

Metodologia para geração de registro de eventos artificial a partir da simulação de eventos discretos.

Trabalho de Conclusão de Curso apresentado a Faculdade de Engenharia da Universidade Federal de Juiz de Fora, como requisito parcial para a obtenção do título de Engenheira de Produção.

Aprovada em 11 de novembro de 2020.

BANCA EXAMINADORA

D.Se., Antônio Ângelo Missiaggia Picorone

Universidade Federal de Juiz de Fora

D.Sc., Fernando Marques de Almeida Nogueira

Universidade Federal de Juiz de Fora

D.Sc., Roberto-Malheiros Moreira Filho

Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Depois de 7 longos, e muito bem aproveitados, anos, aqui está o produto de tantos estudos, risadas, momentos de irritação e questionamentos.

Agradeço minha família: minha mãe, Rosária, meu pai, Marcos e minha irmã Letícia, por esperar esse resultado pacientemente, apesar de sempre perguntarem quando viria. Agradeço especialmente as primas pelo apoio motivacional para pegar o diploma ainda na segunda década do século XXI.

A todos amigos que acompanharam toda a jornada da faculdade: o grupo micro/rpgzada por me tirarem de casa, por fazerem reuniões periódicas e pelas horas que passamos jogando, afinal a vida também precisa isso. O grupo da Marco Zero, porque combater opressões e lutar por um país melhor foi muito mais fácil com vocês por perto. As amizades do intercâmbio que levo para a vida. Agradecimento especial ao grupo Eng que Deu Certo que sempre me rende inúmeras gargalhadas e por compartilharmos tantos momentos de faculdade nesses anos e depois dela também.

Agradecimento especial ao Pedro Uchôas por todos esses anos de amizade sincera desde o ensino fundamental e ao Arthur Avelar pela irmandade e cumplicidade criadas na faculdade, mas que irão para a vida.

Ao meu orientador, Antônio Picorone, que apesar de rigoroso, me ensinou muito sobre o mundo acadêmico e disponibilizou várias horas para me ajudar.

Ao professor Fernando Nogueira pela ajuda com o programa Arena Simulation.

As professoras e professores, que fazem jus à classe, e me engrandeceram imensamente como profissional, pessoa e cidadã.

A universidade pública, gratuita e de qualidade que me permitiu agregar tantas oportunidades, experiências e conhecimentos que eu tenho hoje. Obrigada, UFJF.

Agradeço também aos colegas da SEEPRO, Atlética da Engenharia, Grupo de Educação Tutorial, Diretório Acadêmico, por todo conhecimento e experiências compartilhadas.

RESUMO

A mineração de processos (MP) é uma ferramenta que surgiu juntamente com a indústria 4.0 e vem sendo facilitada pela grande disponibilidade de dados e informações nos dispositivos e softwares. Essa ferramenta faz ponte entre a mineração de dados e a análise de modelos de processos e, com ela, é possível descobrir o modelo de processos a partir de um registro de eventos (RE). O RE é uma entrada importante para a MP e contém informações sobre a execução de cada atividade em um determinado processo. Apesar da indústria 4.0 ter potencializado a criação e armazenamento de dados, nem sempre um RE é fácil de ser obtido, seja por falta da possibilidade do armazenamento desses dados gerados pelo software ou pela máquina, ou pela necessidade de inúmeras autorizações para compartilhamento e uso do mesmo. O objetivo desse trabalho é apresentar uma metodologia capaz de gerar um RE artificial a partir da simulação de eventos discretos de modelos de processos previamente estruturados. Primeiramente, foi construído um modelo de processos usado como base para montar a simulação. O modelo foi replicado e configurado no programa de simulação de eventos discretos de forma a gerar um RE com 6.227 eventos e 812 casos e as informações necessárias para a MP. O RE artificial produzido foi formatado para ser minerado e então filtrado através do Simple Heuristics. A partir da aplicação da MP por meio do algoritmo alpha++, constatou-se uma alta semelhança entre o modelo projetado e o gerado pela MP, o que sugere a factibilidade do uso da metodologia proposta para gerar RE artificiais e servir de auxilio no estudo dos algoritmos de MP.

Palavras-chave: mineração de processos, registro de eventos, modelo de processos, simulação de eventos discretos.

ABSTRACT

Process Mining (PM) has emerged together with the industry 4.0 and it has been promoted thank to the huge amount of data and information available in devices and software. This tool provides a bridge between data mining and business process modeling and analysis, and with it, it is possible to discover the process model from an event log (EL). An EL is an important input to PM and it contains information about the execution of each task in a process. Even though the industry 4.0 has empowered the creation and storage of data, obtaining a EL is not an easy task, either due to the lack of storage ability of all data created in most of software or machines or due to the enormous amount of bureaucracy wrapped to get, share and use the data obtained. This work objective is to present a methodology capable to generate an artificial EL from a discrete-event simulation of a process model designed. Firstly, it was built a process model used as a base to produce the simulation. The model was reproduced and set up in a discrete-event simulation software generating an EL with 6,277 events and 812 cases and containing the information required to PM. The artificial EL produced was, then, formatted and filtered by the Simple Heuristics filter. With the application of the *alpha*++ algorithm, it was observed a great similarity between the process model designed and the one mined from the EL, which suggests a feasibility of the methodology proposed to generate an artificial EL and to serve as a support in PM algorithm studies.

Keywords: process mining, event log, process model, discrete-event simulation.

LISTA DE FIGURAS

Figura 1 - Fragmento de um RE: cada linha é um evento. A coluna Resource
representa o recurso usado nos eventos, no caso, colaboradores. A primeira coluna, Case ID
refere-se aos casos
Figura 2 – Exemplo de processo de negócio usando notação BPMN 20
Figura 3 – Exemplo de modelagem de processo usando a notação de rede de Petr
Figura 4 – Critérios de qualidade para avaliação de modelos de processo 26
Figura 5 – Exemplo de um modelo flor que permite qualquer tipo de sequência de
atividades
Figura 6 - Três tipos básico de MP explicados de acordo com entradas e saídas a
descobrimento, b) conformidade e c) aprimoramento
Figura 7 - Posicionamento dos três tipos principais de MP: descobrimento
conformidade e aprimoramento
Figura 8 – Estrutura de ensaio proposta
Figura 9 – Exemplo da conversão de modelos BPMN-Arena 34
Figura 10 – Adição de módulos para RE
Figura 11 – Diagrama de relacionamento dos blocos Assign e ReadWrite para
criação do registro de eventos
Figura 12 – Processo utilizado para testar o RE artificial
Figura 13 – Número de ocorrência por processo
Figura 14 – Resultado da MP usando o algoritmo <i>alpha</i> ++

LISTA DE TABELAS

Tabela 1 – Fragmento de um RE: cada linha corresponde a um evento
Tabela 2 – Algoritmos utilizados na MP
Tabela 3 – Comparação dos módulos do Arena com o BPMN
Tabela 4 – RE simulado a partir do Arena
Tabela 5 – Configuração do módulo Assign
Tabela 6 – Configuração do módulo <i>ReadWrite</i>
Tabela 7 – Parametrização do módulo <i>File</i>
Tabela 8 – RE resultado da simulação no Arena
Tabela 9 – Distribuição de probabilidade dos processos do modelo
Tabela 10 – Parte do registro de eventos gerado pela simulação após tratamento da
coluna Tf

LISTA DE ABREVIATURAS E SIGLAS

B2B Business-to-business

BPM Business Process Management

BPMN Business Process Management Notation

CRM Customer Relationship Management

CSV Comma-separated Value

ERP Enterprise Resource Management

IoT Internet of Things

MP Mineração de Processos

PAIS Process-aware Information Systems

ProM Process Mining Framework 6.9

RE Registro de Eventos

SCADA Supervisory Control and Data Acquisition

TQM Total Quality Management

UFJF Universidade Federal de Juiz de Fora

WFM Workforce Management

SUMÁRIO

1.	NTRODUÇÃO	13
1.1	JUSTIFICATIVA	15
1.2	ESCOPO DO TRABALHO	16
1.3	OBJETIVOS DO TRABALHO	17
1.4	DEFINIÇÃO DA METODOLOGIA	17
1.5	ESTRUTURA DO TRABALHO	18
2. I	MINERAÇÃO DE PROCESSOS – REVISÃO DA LITERATURA_	19
2.1	PROCESSO	21
2.2	MINERAÇÃO DE PROCESSOS	22
2.3	REGISTRO DE EVENTOS (Event log)	23
2.4	CRITÉRIOS DE QUALIDADE	25
2.5	TIPOS DE TÉCNICAS	27
2.6	ALGORITMOS	30
3. 1	DESENVOLVIMENTO	32
3. 1 3.1		
	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVEN	
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVEN	NTOS
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVEN L 33	NTOS 33
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVEN L 33 3.1.1 Conversão de modelos	NTOS 33 35
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVEN L 33 3.1.1 Conversão de modelos	NTOS 33 35 36
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTA 33 8.1.1 Conversão de modelos	NTOS 33 35 36 39
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTA 33 8.1.1 Conversão de modelos	33 35 36 39 40
3.1 ARTIFICIA	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTA 33 8.1.1 Conversão de modelos	33 35 36 39 40 41
3.1 ARTIFICIA 4. 1	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTA 33 8.1.1 Conversão de modelos	33 35 36 39 40 41
3.1 ARTIFICIA 4. 1	METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTA 33 3.1.1 Conversão de modelos	33 35 36 39 40 41 41 42

4	1.5 RESULTADOS	MINERAÇÃO DE PROCESSOS	44
5.	CONCLUSÕES		46
6.	REFERÊNCIAS		47

1. INTRODUÇÃO

O grande aumento de dados em formato digital, decorrente principalmente do processo de digitalização das organizações que está sendo motivado pelos benefícios advindos da quarta revolução industrial, oferece uma oportunidade aos gestores de aumentar o entendimento dos processos consideravelmente, assim como a eficiência de suas operações. (GÖLZER; CATO; AMBERG, 2015). Diversos sistemas computacionais de gerenciamento de operações disponíveis no mercado capturam sistematicamente informações relativas as atividades dos processos que são realizados intra ou interorganizações (VAN DER AALST, WIL M. P., 2016). Tais sistemas computacionais fazem parte dos ditos Sistemas de Informação Cientes dos Processos (PAIS) que registram os eventos que ocorrem em seus processos (DUMAS; VAN DER AALST; TER HOFSTEDE, 2005), tais como, sistemas de Planejamento de Recursos Empresariais (ERP - Enterprise Resource Planning) (R ADDO-TENKORANG; P HELO, 2011) que registra todas as informações, como por exemplo, preenchimento de formulários, alterações de documentos, login, etc, Gestão de Relacionamento com o Cliente (CRM - Customer Relationship Management) (NGAI; XIU; CHAU, 2009) que registra todas as interações com os consumidores, O Empresa para Empresa (B2B - Business to Business) (BAKRI; STEEL; SOAR, 2010) que registra as trocas de mensagens entre as partes, o Gerenciamento de Força de Trabalho (WFM - Workforce Management) (UNERTL et al., 2010) onde o início e a conclusão das atividades são registrados, os Aplicativos de Controle de Supervisão de Aquisição de Dados (SCADA -Supervisory Control and Data Acquisition) (JICHA; PATTON; CHEN, 2016) que registra toda mudança de estado do sistema sob controle, dentre outros sistemas. O registro de eventos gerado por estes sistemas nada mais é do que um conjunto de dados que registra informações de cada atividade feita dentro do processo. O registro de eventos (RE) gerado por um sistema ou máquina contém informações como data, hora, recurso utilizado, a instância do processo que aconteceu a atividade, duração, entre outros atributos relacionados ao evento.

As informações existentes nos RE são raramente utilizadas pelas organizações para agregar valor às suas operações. O que se vê com mais frequência é a busca manual no RE de eventos que precederam algum comportamento anormal no processo monitorado. Esta busca visa identificar uma possível causa do problema observado na operação. Com o objetivo de extrair conhecimentos de uma forma mais inteligente dos RE, oferecendo a oportunidade de entender uma determinada operação a partir dos dados históricos das

atividades realizadas em busca de operações mais eficientes, surge a Mineração de Processos (MP) (WESTERGAARD, 2011).

A MP pode ser vista como uma ferramenta de suporte ao gerenciamento de processos. O gerenciamento de processos vem passando por evoluções constantes desde a década de 60, passando pela Gestão Total da Qualidade (TQM – *Total Quality Management*), reengenharia, Lean Six Sigma e atualmente o Gerenciamento de Processos de Negócios (VAN DER AALST, W.M.P.; WEIJTERS, 2004).

A MP vem se mostrando uma ferramenta extremamente poderosa dentro do gerenciamento de processos, é um tópico de pesquisa relativamente recente que faz uma ponte entre dois tópicos importantes: a mineração de dados e a modelagem de processo (VAN DER AALST, WIL M.P., 2011). O primeiro representa a identificação automática de padrões, anomalias e correlações em grandes conjuntos de dados das atividades realizadas e o segundo é a representação de processos de negócio, com o objetivo de documentar, entender e analisar as atividades. Além de ser uma ponte entre esses dois importantes tópicos, a MP tem como objetivos descobrir, monitorar e melhorar o processo real através das informações extraídas dos RE disponibilizados nos sistemas PAIS. (VAN DER AALST, W.M.P.; WEIJTERS, 2004).

Apenas armazenamento e análise de dados não são suficientes nos dias atuais, fazse necessário relacionar dados com processos operacionais e com a MP, além do descobrimento do modelo de processo que é executado, é possível checar conformidade com o modelo de processo planejado, quantificar e identificar desvios no processo real, encontrar gargalos, prever saídas de processos, identificar relações entre os recursos e várias outras vantagens de se entender o processo a fundo (VAN DER AALST, WIL M. P., 2016). Alinhado aos conhecimentos da Engenharia de Produção, a MP pode ser uma valiosa ferramenta no entendimento e otimização de processos.

Para que seja possível estudar e desenvolver novos algoritmos para uso na MP é de suma importância ter em mãos um RE que pode ser utilizado como ponto de partida do estudo e sua qualidade pode impactar diretamente no estudo da MP (SURIADI *et al.*, 2016). Entretanto, o acesso a um RE com os atributos necessários para a aplicação da MP desejada nem sempre é fácil de se obter, em parte, devido às políticas de *compliance* das empresas e ainda, o fato de que muitas empresas ainda possuem registros de dados dos eventos realizada

de forma manual (IVERS; BYRNE; BYRNE, 2016). A inexistência de um RE adequado inviabiliza o estudo da MP.

1.1 JUSTIFICATIVA

Tendo em vista as oportunidades que vieram da quarta revolução industrial, especialmente com o advento da Internet das Coisas (IoT - Internet of Things), grandes conjuntos de dados provenientes dos processos industriais passaram a ser acessíveis. Estes dados podem ser encontrados em diversas fontes como aparelhos, softwares, máquinas, lugares, conteúdos e interações que envolvam quaisquer dispositivos inteligentes. Extrair conhecimento desse conjunto de dados para suportar o processo decisório nas empresas passa a ser um grande diferencial competitivo e objeto de desejo dos gestores de processos. Mas nem sempre é possível obter um RE que registre todos e quaisquer eventos realizados pela máquina, programa ou usuário, ou que tenha as características necessárias para a MP. A falta de um RE estruturado com as características necessárias para a MP pode inviabilizar o estudo dessa ferramenta que tem um potencial surpreendente para a análise, descobrimento, comparação e melhoria de modelos de processos. O desenvolvimento de um RE artificial, que pode ser usado seguramente para realizar a MP, abre portas e diminui barreiras para o estudo de modelo de processos, bem como facilita a manipulação dos fatores do modelo de processo a serem estudados e analisados.

A Engenharia de Produção está diretamente envolvida com os sistemas produtivos, principalmente no que tange a operação e a melhoria desses sistemas ("ABEPRO - Associação Brasileira de Engenharia de Produção | A Profissão", [S.d.]). A união da Engenharia de Produção com a MP pode ser considerada uma união providencial, pois reúne uma ferramenta que trata grandes conjuntos de dados provenientes de eventos gerados nos processos, com a necessidade de se extrair conhecimentos destes dados que podem ser cruciais para tomadas de decisão, entendimento das análises e o impacto que as ações podem refletir no processo como um todo.

Dessa forma, este trabalho se justifica pela importância de se facilitar o acesso à um *input* indispensável para o estudo da MP, o RE, dando flexibilidade ao pesquisador de MP para buscar soluções e ferramentas que podem trazer um diferencial competitivo às empresas quanto a gestão de seus processos.

Ademais, mesmo que a MP tenha se mostrada como uma poderosa ferramenta para auxiliar na gestão dos mais diversos tipos de processos, tema altamente aderente a grande área da Engenharia de Produção, pouco se tem discutido sobre o tema na área e, em especial, na Universidade Federal de Juiz de Fora (UFJF). Uma vez que, até então, nenhum trabalho de graduação na UFJF tenha sido apresentado explorando tal tema, este trabalho se justifica também pelo ineditismo do tema no curso de Engenharia de Produção da UFJF. Assim, poderá servir como instrumento de provocação e estímulo às discussões e pesquisas sobre a aplicação de ferramentas emergentes que auxiliarão os atuais e futuros engenheiros de produção a levar as empresas rumo à Indústria 4.0.

1.2 ESCOPO DO TRABALHO

O trabalho envolve a produção e análise de um RE artificial a partir de um único modelo de processos previamente estruturado. Tal modelo foi usado na construção de uma simulação que registrou todos seus eventos executados em um arquivo externo, o RE artificial, que foi sujeitado então à MP para validar a relação do modelo resultante com o modelo inicialmente proposto.

A questão central do trabalho é a criação de um RE a partir de um modelo de processo previamente escolhido, de forma a facilitar a obtenção do principal *input* para a MP, visto as dificuldades de se obter tais dados, fundamentais para o estudo da MP.

Este trabalho se limita ao uso de um único modelo de processo hipotético e simples, com apenas 9 tarefas, imutáveis, exclusivamente para o teste do RE artificial desenvolvido. O programa de simulação de eventos discretos utilizado nesse trabalho, o *Arena Simulation*, em seu módulo para estudantes, só pode processar 150 entidades concomitantemente, o que limitou não só o tamanho do modelo de processos, mas também os parâmetros utilizados, evitando gargalos na simulação.

Para a MP foi utilizado o programa *Process Mining Framework 6.9* e o algoritmo *alpha*++, visto que não precisa de parâmetros iniciais para a MP, o que diminui as variações de resultado que a troca de parâmetros poderia causar ao resultado final.

1.3 OBJETIVOS DO TRABALHO

Este trabalho visa diminuir a lacuna existente na obtenção de RE, propondo o uso de ferramentas computacionais de simulação para gerar RE com os atributos e quantidade de eventos necessários para ser utilizado em pesquisas de MP.

Neste sentido, o principal objetivo deste trabalho é a apresentação de uma metodologia para geração de um RE artificial a partir de uma ferramenta computacional de simulação baseada em eventos discretos.

É um objetivo secundário deste trabalho a introdução do tema dentro da Engenharia de Produção da Universidade Federal de Juiz de Fora, visto que o tema ainda não foi abordado em nenhum trabalho de conclusão de curso até esta data. O tema pode ser amplamente explorado por essa área de conhecimento tendo em vista as oportunidades de estudo e aplicação sobre MP e o entendimento sobre processos e suas otimizações que são feitos por Engenheiros e Engenheiras de Produção.

1.4 DEFINIÇÃO DA METODOLOGIA

O propósito do presente estudo pode ser classificado como **exploratório**, uma vez que procura proporcionar maior familiaridade com os conceitos ligados a MP e explicitá-los de forma a entender o tema. Para isso foi feito um levantamento bibliográfico dos tópicos que o tangencia e que são necessários para melhor entendimento com relação a essa ferramenta.

A abordagem é **quantitativa** e conta com análise de resultados a partir dos dados de entrada, mais especificamente do RE artificial gerado pelo programa de simulação de eventos discretos e, em seguida, pela MP com as informações do modelo gerado, resultando em informações e atributos mensuráveis do processo que podem ser analisados e comparados com o usado para dar início à simulação.

Em busca de atingir o objetivo da Seção 1.3, foram realizados os seguintes passos:

 Foi definido um escopo de atuação para limitar o tema e o objetivo de pesquisa, demarcando os principais conceitos e quais tópicos seriam aprofundados;

- 2. Foi realizada uma ampla revisão da literatura sobre a MP. A pesquisa sobre o tema foi realizada em grande parte através de livros e artigos, mas também através de um curso online específico sobre MP¹.
- 3. Foi escolhido um programa de simulação de eventos discretos que permitisse registrar as informações das iterações suficientes para produzir um RE e que fosse largamente usado na Engenharia de Produção;
- 4. Foi escolhido um modelo de processo que pudesse ser usado como referência no estudo e que envolvesse características genéricas, comumente encontradas na prática, tais como, paralelismo e loops. O modelo foi então replicado no programa de simulação de eventos e parametrizado com informações como a taxa de entrada, distribuição de probabilidade dos eventos e das escolhas. Também foi configurado o arquivo de saída da simulação que resultou no RE;
- 5. Foi conduzido um estudo para conhecer as opções de ferramentas computacionais (softwares para MP) disponíveis no mercado, definição das características desejadas da ferramenta para a análise do RE escolhido e, finalmente, a escolha da ferramenta a ser utilizada. A escolha foi feita considerando artigos científicos, estudos dos critérios de qualidade dos artigos e facilidade de acesso ao programa;
- 6. A partir da produção do RE e da escolha da ferramenta computacional realizadas nas etapas "3" e "4", respectivamente, foi realizada a MP do RE;
- 7. Em seguida, foi comparado os parâmetros do modelo de processo de entrada no programa de simulação com o resultado, não só do modelo de processos da MP, como também nas probabilidades apresentadas;
- 8. Finalmente, com os resultados propriamente analisados, foi realizada uma análise crítica do RE artificial, considerando toda sua criação e comparação com o resultado obtido, auto avaliação também com relação ao tema e o trabalho realizado.

1.5 ESTRUTURA DO TRABALHO

O presente trabalho está dividido em cinco capítulos, cujo conteúdo é o que se segue:

¹ Curso Online "Mineração de Processos: ciência de dados na prática", oferecido pela Eindhoven University of Technology TU/e que pode ser acesso através do site: https://www.coursera.org/learn/processmining.

- 1) Capítulo 1: contém a introdução do estudo que será realizado, a contextualização do tema, a apresentação dos termos que serão amplamente utilizados, o problema que levou à realização do estudo bem como a importância do tema e a justificativa para sua realização, escopo que o trabalho pretende abordar, metodologia e sua estrutura de divisão;
- Capítulo 2: apresenta a revisão da literatura que o estudo é embasado: A MP e suas principais técnicas, ferramentas e conceitos fundamentais para o melhor entendimento do assunto que serão discutidos;
- 3) Capítulo 3: apresenta as etapas para a criação do RE artificial a partir de um modelo de processo de negócio BPMN (Business Process Model and Notation) e os resultados obtidos quando se submete o RE artificial gerado pela simulação um modelo de processo previamente escolhido à técnica de descobrimento na MP.
- 4) Capítulo 4: análise dos resultados da MP e comparação com o modelo do processo de entrada, considerando os parâmetros utilizados;
- Capítulo 5: serão abordadas as considerações finais a respeito do trabalho desenvolvido, assim como as dificuldades, limitações e recomendações para outras possíveis pesquisas;

2. MINERAÇÃO DE PROCESSOS – REVISÃO DA LITERATURA

Esse capítulo descreve os principais tópicos necessários para que o tema seja entendido com clareza. Inicialmente será abordada a definição de processos. Em seguida, será apresentada uma revisão bibliográfica sobre os registros de eventos. Finalmente, uma revisão sobre a MP, seu histórico, algoritmos e programas que podem ser utilizados serão discutidos.

Antes da revisão da literatura, é importante rever alguns conceitos que serão utilizados com frequência no contexto de MP. São eles (VAN DER AALST, WIL M. P.; VAN HEE, 2002):

- a) <u>Evento</u>: unidade de trabalho lógica feita inteiramente por um recurso. Pode ser referenciado também como tarefa ou atividade;
- b) <u>Recurso</u>: é um nome genérico que pode se referir a uma ou mais pessoas ou máquinas que executam uma tarefa específica. O recurso não necessariamente é o executor da tarefa, mas é o responsável pela execução;
- c) <u>Caso</u>: é a instância produzida ou transformada num trabalho. Pode ser um produto, serviço, item. Cada caso deve ter sua própria identidade. O fragmento de um RE apresentado na Figura 1 exemplifica uma das formas como os conceitos apresentados de evento, recurso e caso podem aparecer;

Case Id	Attributes			
	Activity	Timestamp	Resource	Location
1	Present at ED	2014-10-03 07:54	Jason	Emergency
1	Triage request	2014-10-03 07:57	Jason	Emergency
1	Triage	2014-10-03 08:03	Susan	Emergency
1	Medical assign	2014-10-03 08:10	Jason	Emergency
1	Blood tests	2014-10-03 09:34	Sarah	Laboratory
1	Admit to hospital	2014-10-03 10:02	George	Ward
2	Present at ED	2014-10-03 08:12	Jason	Emergency
2	Triage request	2014-10-03 08:16	Jason	Emergency
2	Triage	2014-10-03 08:20	Ross	Emergency
2	Medical assign	2014-10-03 08:30	Jason	Emergency
2	Discharge to home	2014-10-03 08:50	Jason	Emergency

Figura 1 – Fragmento de um RE: cada linha é um evento. A coluna *Resource* representa o recurso usado nos eventos, no caso, colaboradores. A primeira coluna, *Case ID*, refere-se aos casos.

Fonte: Suriadi et al. (2016)

d) Modelo de Processo: um procedimento com tarefas e conjuntos de condições que determina a ordem de execução das atividades é um processo. O modelo de processo usando notações BPMN pode ser visto na Figura 2. As setas são chamadas de arcos e mostram a orientação do fluxo. Os losangos com "X" são as transições que classificam escolhas;



Figura 2 – Exemplo de processo de negócio usando notação BPMN Fonte: VAN DER AALST, WIL M.P. (2011b)

e) Rede de Petri: É uma linguagem de modelo de processo permitindo a modelagem de concorrência. Consiste de um grafo direcionado e bipartido² que consiste em lugares e transições, como na Figura 3. No caso da Rede de Petri, os lugares são marcados pelos círculos como c1, c2, c3, c4 e c5 e os quadrados a, b, d, e, h são as transições, ou seja, as atividades a serem executadas no processo depois que suas respectivas transições forem disparadas;

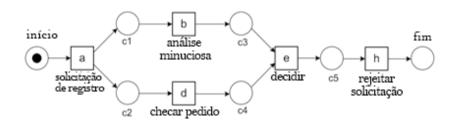


Figura 3 – Exemplo de modelagem de processo usando a notação de rede de Petri Fonte: VAN DER AALST, WIL M.P. (2011b)

f) <u>Traço</u>: sequência completa de eventos dentro de um processo (SONG; GÜNTHER; VAN DER AALST, 2009). Na Figura 3, por exemplo, um traço seria a ordem de execução a-b-d-e-h e um outro traço seria a-d-b-e-h.

2.1 PROCESSO

Antes de entender a MP é preciso definir o que é um processo e porque ele é tão importante. Existem diversas definições de Processo na literatura. De uma forma geral, processo pode ser definido como sendo o meio como as empresas e organizações oferecem produtos, podendo estes serem bens ou serviços, aos seus clientes. Todas as empresas realizam processos (SMART; MADDERN; ROGGER, 2009) e essas atividades são fundamentais para a eficiência da instituição, visto que é o que baseia a entrega final de seus produtos. O objetivo de um processo é a transformação de insumos em produtos com valor agregado, que serão postos à disposição de clientes internos ou externos (BALDAM; VALLE; ROZENFELD, 2014). Com um pouco mais de detalhes, os processos são conjuntos de atividades, executadas de forma coordenada em um ambiente técnico e organizacional (WESKE, 2012), que acontecem em um período de tempo específico com um começo, um

² Um grafo é bipartido quando o seu conjunto de vértices puder ser particionado em dois subconjuntos tais que não haja ligações internas a eles, ou seja, entre dois vértices de um mesmo subconjunto (NETTO; JURKIEWICZ, 2009).

fim, além de entradas (*inputs*) e saídas (*outputs*) claramente identificados, definindo assim uma estrutura para ação (DAVENPORT, 1994). O processo também pode ser visto como um grupo de tarefas logicamente interligadas, que utiliza recursos da organização para gerar resultados definidos, de forma a apoiar os objetivos da organização (HARRINGTON; ESSELING; NIMWEGEN, 1997)

Entender e evidenciar a importância que os processos têm para as organizações é o primeiro passo para compreender a importância de gerencia-los corretamente (FRANCISCO; SANTOS, 2011). Conhecer os processos e suas características é indispensável para identificar oportunidades de melhoria, obter dados para tomar decisão e gerar a base para definir metas e avaliar rotinas e resultados (PAMPONET, 2019).

2.2 MINERAÇÃO DE PROCESSOS

O termo mineração de processos surgiu em 1999 e foi utilizado para se referir à metodologia de extração de uma descrição estruturada de processo a partir de um conjunto de execuções reais (LOPES, 2017). A MP tem como objetivo a construção automática de modelos de processo explicando o comportamento observado em um RE. Essa técnica complementa uma abordagem já existente, o BPM (business process management) ou simplesmente Gestão de Processos do Negócio, abordagem que combina os fundamentos vindos da tecnologia de informação e da gestão de conhecimento aplicado em um processo operacional de negócios (VAN DER AALST, W. M. P. et al., 2007).

A gestão de processos do negócio é uma área de conhecimento que junta os conceitos, metodologias e atividades envolvidas no apoio ao ciclo de vida de um processo. Essa disciplina trata gerencialmente processos de negócios como ativo das organizações e sua premissa é que através da transformação, controle, definição e desenho contínuo dos processos seja possível alcançar os objetivos organizacionais (LOPES, 2017).

O fluxo de trabalho é um conjunto de atividades com um objetivo em comum e envolve diversas atividades ordenadas a serem executadas por humanos ou outro recurso da empresa afim de atingir um requerimento pré-definido. A automatização do fluxo de trabalho permite às organizações se tornarem mais flexíveis e adaptáveis através da contínua customização do trabalho para conseguirem acompanhar as mudanças do ambiente empresarial (SALIMIFARD; WRIGHT, 2001).

É possível comparar o BPM com o Fluxo de Trabalho. O fluxo de trabalho foca na automação dos processos do negócio, enquanto o BPM foca um escopo mais amplo: utiliza a automação e a análise de processos para fazer a gestão e organização dos processos e do trabalho (VAN DER AALST, WIL M.P., 2011).

De forma geral, a MP permite a extração de dados dos processos que são encontrados em um ambiente tecnológico e, através deles, constatar informações sobre o fluxo de atividade real do processo (fluxo de trabalho) e da rede social onde o processo se encontra (FRANCISCO; SANTOS, 2011).

2.3 REGISTRO DE EVENTOS (EVENT LOG)

Os sistemas de informação cientes de processos (PAIS - process-aware information systems) são sistemas que armazenam informações sobre a ocorrência de eventos em um processo (JUNIOR, 2017). Alguns PAISs são bem conhecidos como o ERP, CRM, SCADA etc., e armazenam informações detalhadas sobre os eventos de forma estruturada. Esses sistemas podem ser usados como base para a MP.

O armazenamento das atividades e das informações é de fundamental importância para que a MP possa ser executada. Isso porque a base para a mineração são os registros de eventos (*event logs*).

Normalmente esses registros contém informações sobre o processo referenciado com as atividades e as instâncias que estão sendo observadas. Em geral, a instância trata especificamente o caso que está sendo manipulado, podendo ser uma ordem de cliente, candidatura de trabalho, reivindicação de seguro, etc (VAN DER AALST, WIL M.P., 2011).

Na Tabela 1 é possível observar um exemplo de informações que podem ser armazenadas em um RE a partir da Figura 3³. Cada linha representa um evento e os eventos estão agrupados por instância, como pode ser visto na primeira coluna da Tabela 1 (ID Caso). Cada evento tem seu registro próprio, justamente para diferenciar eventos iguais em instâncias diferentes. Além disso, também é registrado o horário que o evento foi realizado, assim como os responsáveis pela atividade e o custo.

_

³ Tal exemplo pode ser encontrado no site www.processmining.org.

ID Caso	Propriedades				
ID Cast	ID do Evento	dd-MM-aaaa:HH.mm	Atividade	Recurso	Custo
1	35654423	30-12-2010:11.02	Solicitação de registro	Pete	50
	35654424	31-12-2010:10.06	Análise minuciosa	Sue	400
	35654425	05-01-2011:15.12	Checar pedido	Mike	100
	35654426	06-01-2011:11.18	Decidir	Sara	200
	35654427	07-01-2011:14.24	Rejeitar solicitação	Pete	200
2	35654483	30-12-2010:11.32	Solicitação de registro	Mike	50
	35654485	30-2-2010:12.12	Checar pedido	Mike	100
	35654487	30-12-2010:14.16	Examinar	Sean	400
	35654488	05-01-2011:11.22	Decidir	Sara	200
	35654489	08-01-2011:12.05	Pagar compensação	Ellen	200
3	35654521	30-12-2010:14.32	Solicitação de registro	Pete	50
	35654522	30-12-2010:15.06	Examinar	Mike	400
	35654524	30-12-2010:16.34	Checar pedido	Ellen	100
	35654525	06-01-2011:09.18	Decidir	Sara	200
	35654526	06-01-2011:12.18	Reiniciar solicitação	Sara	200
	35654527	06-01-2011:13.06	Análise minuciosa	Sean	400
	35654530	08-01-2011:11.43	Checar pedido	Pete	100
	35654531	09-01-2011:09.55	Decidir	Sara	200
	35654533	15-01-2011:10.45	Pagar compensação	Ellen	200
4	35654641	06-01-2011:15.02	Solicitação de registro	Pete	50
	35654643	07-01-2011:12.06	Checar pedido	Mike	100
	35654644	08-01-2011:14.43	Análise minuciosa	Sean	400
	35654645	09-01-2011:12.02	Decidir	Sara	200
	35654647	12-01-2011:15.44	Rejeitar solicitação	Ellen	200
5	35654711	06-01-2011:09.02	Solicitação de registro	Ellen	50
	35654712	07-01-2011:10.16	Examinar	Mike	400
	35654714	08-01-2011:11.22	Checar pedido	Pete	100
	35654715	10-01-2011:13.28	Decidir	Sara	200
	35654721	20-01-2011:12.48	Reiniciar solicitação	Sara	200
	35654725	23-01-2011:13.12	Decidir	Sara	200
	35654726	24-01-2011:14.56	Rejeitar solicitação	Mike	200
6	35654871	06-01-2011:15.02	Solicitação de registro	Mike	50
	35654873	06-01-2011:16.06	Examinar	Ellen	400
	35654874	07-01-2011:16.22	Checar pedido	Mike	100
	35654875	07-01-2011:16.52	Decidir	Sara	200
	35654877	16-01-2011:11.47	Pagar compensação	Mike	200

Tabela 1 – Fragmento de um RE: cada linha corresponde a um evento. Fonte: Adaptado de VAN DER AALST, WIL M.P. (2011b)

É importante ressaltar que o custo, observado na última coluna da Tabela 1 – Fragmento de um RE: cada linha corresponde a um evento. pode ser substituído por qualquer tipo de atributo de dados. Nem todos os eventos têm o mesmo conjunto de atributos, mas

tipicamente, todos os eventos relativos à uma mesma atividade, tem o mesmo conjunto de atributos (VAN DER AALST, WIL M.P., 2011).

Os registros de eventos devem assumir as seguintes premissas (VAN DER AALST, WIL M.P., 2011):

- a) Cada RE está conectado à um único processo. É preciso ter a certeza que todas as tarefas realizadas fazem parte daquele modelo de processo específico a ser realizado;
- b) Cada evento deve se referir a uma atividade, uma etapa definida do processo, não podendo referir-se a um determinado período de tempo. Exemplo de uma atividade: seu começo e fim, e não sua duração (registrado na Tabela 1 como coluna "atividade");
- c) As atividades devem possuir um identificador único (que na Tabela 1 é registrado na coluna "ID do Evento") e devem referir-se à apenas uma atividade;
- d) Todo evento no registro deve estar referenciado à uma única instância, ou seja,
 à um único caso;
- e) Os eventos dentro de um caso precisam estar ordenados. Sem a ordenação é impossível descobrir a relação de dependência no modelo de processo.

2.4 CRITÉRIOS DE QUALIDADE

RE nem sempre estão completos, isso porque o registro só mostra atividades que já aconteceram no sistema. Ou seja, ainda podem, em alguns casos, existir uma infinidade de combinações que ainda não aconteceram (VAN DER AALST, W. M. P. *et al.*, 2007).

Além disso, existem dois fenômenos relacionados que podem fazer com que o RE seja menos representativo para o processo que está sendo analisado: o ruído e a incompletude.

O ruído acontece quando os registros de eventos contêm comportamentos raros e infrequentes que não representam o comportamento típico do processo. É impossível para o algoritmo distinguir um erro de um comportamento excepcional, para isso é necessária uma avaliação humana externa do processo. Com relação ao ruído, é possível considerar o modelo 80/20. Na maioria dos casos será interessante que o processo descreva 80% dos comportamentos que podem ser vistos no registro.

A incompletude, por outro lado, acontece quando os registros contém apenas uma pequena porção dos comportamentos possíveis, talvez graças à um grande número de alternativas (VAN DER AALST, WIL M.P., 2011).

Enquanto o ruído trata de "informações de mais", ou seja, desvios de comportamento que podem não representar o que acontece no processo, a incompletude trata de "informação de menos", informações que representam apenas uma parte do que pode acontecer no processo e deixam muitos comportamentos de fora do registro. Isso acontece porque não é possível assumir que todo o RE contém todos os possíveis comportamentos para aquele processo. Modelos de processos, tradicionalmente, permitem um número exponencial, e até mesmo infinito, de traços (em casos de loops) (VAN DER AALST, WIL M.P., 2011).

Os dois critérios de qualidade que acabaram de ser citados se referem à qualidade do RE, ou seja, a entrada para a MP. Existem critérios de qualidade também para o modelo produzido, a saída da MP. A qualidade do modelo produzido é avaliada segundo os critérios de ajuste, simplicidade, precisão e generalização (Figura 4) (WESTERGAARD, 2011).



Figura 4 – Critérios de qualidade para avaliação de modelos de processo Fonte: Adaptado de GANDULFO (2016)

Um modelo com um bom **ajuste** permite o máximo de comportamentos visto no registro. O modelo tem um ajuste perfeito se todos os traços no registro podem ser replicados pelo modelo de início ao fim.

No contexto de MP, a **simplicidade** refere-se à quantidade de transições e arcos existentes em um modelo. O modelo que melhor otimizar o uso de arcos e transições e ainda assim representar o RE é o modelo mais simples (WESTERGAARD, 2011).

Mas um modelo que só pode replicar os exemplos mostrados no registro não é um bom modelo, pois os comportamentos apresentados no registro são amostrais e não apresentam todos os comportamentos possíveis. Um modelo é considerado **preciso** se não

permite comportamento além do registrado pelo RE, comportamento que pode destoar completamente da realidade do processo. Um modelo que não é preciso é *underfitting*, ou seja, é *subajustado* e generaliza o comportamento no RE. Por exemplo, se um modelo permite comportamentos muito diferentes daqueles que podem ser observados no registro, pode-se dizer que não é preciso e que existe um subajuste. Para melhor compreensão, na Figura 5 é ilustrado um modelo flor. Esse modelo mostra todas as atividades que existem no registro e mais nenhum tipo de conhecimento pode ser retirado de tal modelo, visto que não tem ordenação, simultaneidade, escolhas, e outros indicativos de um modelo de processo convencional. O modelo flor carece precisão. E por isso a necessidade de critérios além da simplicidade e ajuste (VAN DER AALST, WIL M.P., 2011).

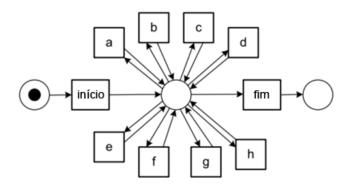


Figura 5 – Exemplo de um modelo flor que permite qualquer tipo de sequência de atividades Fonte: VAN DER AALST, WIL M.P. (2011)

Um modelo também deve **generalizar** e não se restringir aos comportamentos vistos no registro. Um modelo que não generaliza é *overfitting*, que trataremos aqui como *sobreajustado*. O sobreajuste é o problema que ocorre quando um modelo muito específico é gerado. Isso significa que ele é capaz de replicar um modelo do registro, mas qualquer outro que venha a surgir ele não consegue absorver no modelo proposto.

Os algoritmos de MP precisam balancear entre o subajuste e o sobreajuste. Um modelo com muito sobreajuste só permite o exato comportamento do registro e um modelo muito subajustado deixa de lado as características do processo e generaliza demais o modelo, perdendo as características de tomada de decisão e escolhas que o processo venha a ter (WESTERGAARD, 2011).

2.5 TIPOS DE TÉCNICAS

Assumindo que é possível registrar os eventos que acontecem em um determinado sistema na sequência em que ocorrem e que cada um deles refere-se a uma atividade, e que estão agrupados por instâncias que ocorrem, como é o caso do registro representado na Tabela 1, pode-se usar tal informação para conduzir três tipos diferentes de MP: o descobrimento, conformidade e o aprimoramento. Observa-se na Figura 6 a diferença entre os tipos de MP e evidencia as entradas e saídas de cada uma delas.

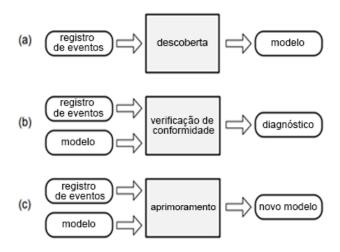


Figura 6 – Três tipos básico de MP explicados de acordo com entradas e saídas a) descobrimento, b) conformidade e c) aprimoramento

Fonte: Adaptado de VAN DER AALST, WIL M.P. (2011b)

Pode-se observar na Figura 7 as técnicas de MP e estabelece uma conexão entre o processo real, juntamente com suas informações, e o modelo do processo nos três tipos de MP:

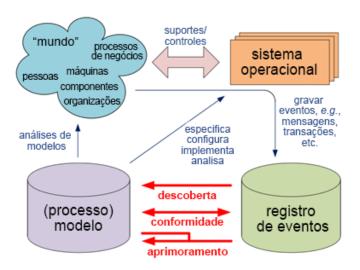


Figura 7 – Posicionamento dos três tipos principais de MP: descobrimento, conformidade e aprimoramento

Fonte: VAN DER AALST, WIL M.P. (2011)

- a) Descobrimento: A técnica utiliza o RE e produz um modelo sem nenhuma informação a-priori ou conhecimento adicional. Essa técnica é a mais proeminente dentre as três. A entrada para tal técnica é um RE e a saída é a produção de um modelo de processo. Além disso, dependendo do nível de informações adicionais no RE, é possível ainda descobrir o modelo de relação entre os recursos (por exemplo, como estão divididas as funções da organização). (VAN DER AALST, WIL M.P., 2011).
- b) Conformidade: nesse caso, considera-se a existência de um modelo de processo para ser comparado com o RE do mesmo processo. A checagem de conformidade pode ser usada para comparar a realidade, como gravado no RE, e o modelo preliminar e vice-versa. Nesse caso, vários tipos de perspectivas podem ser utilizados para comprar a realidade com o modelo criado. A técnica da conformidade mede os alinhamentos entre o modelo e a realidade. A conformidade necessita, além do RE, de um modelo de processos como entrada. A saída dessa técnica consiste em diagnósticos mostrando as diferenças e similaridades entre o modelo e o RE (WESTERGAARD, 2011). A conformidade, no geral, é usada para detectar, localizar, quantificar e explicar desvios e para medir a severidade desses desvios (VAN DER AALST, WIL M.P., 2011).
- c) Aprimoramento: essa técnica objetiva a expansão e o melhoramento de um modelo de processo existente usando informações sobre o processo como gravado no RE. O aprimoramento procura modificar ou estender o modelo apriori. Assim como para a conformidade, é necessário o registro e o modelo do processo como entrada (WESTERGAARD, 2011). Um tipo de aprimoramento é o conserto, quando é feita uma modificação no modelo para melhor refletir a realidade. Outro tipo é a extensão, usada para adicionar perspectivas no modelo do processo através da correlação com o registro (VAN DER AALST, WIL M.P., 2011).

Registros de eventos como o exemplo mostrado anteriormente podem ser usados para descobrir qual o modelo de processo envolvido. Mas a MP não está limitada apenas ao descobrimento do fluxo das atividades, essa é apenas um dos tipos de perspectivas envolvidas na ferramenta. A MP envolve quatro tipos de perspectivas: o controle de fluxo, organizacional, temporal e, por último, perspectiva por instância. A perspectiva do controle de fluxo, a qual, inclusive, esse trabalho está limitado, tem como objetivo achar uma boa caracterização para todos possíveis rastros que podem ocorrer. A perspectiva organizacional foca na informação sobre os recursos escondidos no registro, ou seja, como pessoas, funções, sistemas, departamentos estão envolvidos e como estão relacionados, objetivando estruturar a organização através da classificação dos recursos utilizados. A perspectiva por instância foca em como os casos são caracterizados pelos seus valores dos dados correspondentes (por exemplo, se um caso representa uma ordem de reabastecimento, pode ser interessante saber o fornecedor e o número de produtos pedido). Por último, existe a perspectiva temporal que foca na cronometragem e na frequência dos eventos. Através do carimbo de hora pode-se descobrir gargalos, níveis de serviço, monitorar utilização de recursos, prever tempo remanescente de processo, etc. (VAN DER AALST, WIL M.P., 2011).

É importante notar que as diferentes perspectivas são parcialmente sobrepostas e não-exaustivas. Ainda assim elas provêm uma boa caracterização dos aspectos da MP no que tange a análise.

2.6 ALGORITMOS

Para este trabalho, apenas os algoritmos de descobrimento serão abordados, isso porque a criação e descobrimento do modelo de processo a partir da mineração são o foco principal desse trabalho. Com este intuito, vários algoritmos podem ser usados na MP no que tange a técnica descobrimento, usados para gerar modelos de processos a partir de um RE (JUNIOR, 2017).

O resultado da MP pode diferir dependendo do algoritmo utilizado e a escolha também pode interferir nos critérios de qualidade. Na tabela 2 alguns algoritmos utilizados na MP estão listados e descritos.

Algoritmo	Descrição
α Miner	Descobrir uma rede de Petri. Um dos primeiros algoritmos
	capazes de lidar com a concorrência. O algoritmo ganhou

Algoritmo	Descrição	
	atualizações que otimizaram suas limitações como alpha+	
	(DE MEDEIROS et al., 2004) e o alpha++ (WEN; VAN	
	DER AALST; et al., 2007).	
Heuristics Miner	Descobrir uma rede heurística rápida usando uma abordagem	
	probabilística. Considerando o ruído, os resultados desse	
	algoritmo estão alinhados com o processo.	
Inductive Miner	Descobrir uma árvore de processo ou uma rede de Petri.	
	Considerando os comportamentos infrequentes e grandes	
	registros de eventos, o algoritmo gera um bom resultado e tem	
	uma precisão garantida.	
Fuzzy Miner	Descobrir um modelo fuzzy. Excelente algoritmo no que	
	tange comportamentos não-estruturados. Além disso é rápido	
	e completa todos os rastros do processo.	
Genetic Miner	Descobrir uma rede heurística usando algoritmo genético.	
	Esse algoritmo depende das configurações do parâmetro para	
	que seu resultado possa ser considerado confiável e eficiente.	
ILP Miner	Descobrir rede de Petri resolvendo problemas de	
	Programação Linear Inteira (ILP). Os resultados tendem a	
	apresentar uma perfeita capacidade, mas no geral é fraco na	
	precisão e, além disso, não apresenta eficiência quando	
	processa grandes registros de evento.	

Tabela 2 – Algoritmos utilizados na MP Fonte: Adaptado de Aguirre et al. (2019)

O desafio dos algoritmos é balancear os quatro critérios de qualidade. Por exemplo, um modelo muito simples tem mais chances de ser subajustado ou ter menos precisão.

O algoritmo *alpha* é um tipo de algoritmo determinístico, ou seja, faz parte dos algoritmos que produzem resultados definidos e reprodutíveis (JUNIOR, 2017). São algoritmos que são capazes de reproduzirem os comportamentos observados pelo RE. Esse algoritmo foi um dos primeiros desenvolvidos na técnica de descobrimento da MP e, apesar da fácil aplicação, possui algumas limitações como: problemas para lidar com dependências implícitas, loops curtos e longos, ruídos no RE e escolhas não-livres. Com isso vieram as modificações *alpha*+ e *alpha*++ (WEN; VAN DER AALST; *et al.*, 2007).

A primeira modificação, *alpha+*, foca na melhoria de compreensão de pequenos loops e a segunda melhoria, alpha++, aproveitando as modificações feitas por sua anterior, insere também a melhoria de escolhas não-livres e consegue identificar dependências implícitas e explícitas entre as atividades (EFFENDI; SARNO, 2020). Ambas as melhorias ainda tem limitações como dificuldade de lidar com ruído, mas isso pode ser contornado através de filtragem de comportamentos menos frequentes (WEN; VAN DER AALST; *et al.*, 2007).

O uso do algoritmo *alpha*++ nesse trabalho se dá pela facilidade de uso do algoritmo, uma vez que ele não depende de parâmetros iniciais para executar a mineração, evitando assim diferenças no modelo final minerado e pela aplicabilidade em pequenos RE artificiais, apresentando notáveis melhorias com relação ao seu gerador, *alpha miner* (WEN; VAN DER AALST; *et al.*, 2007).

3. DESENVOLVIMENTO

É representado na Figura 8 – Estrutura de ensaio proposta o diagrama em blocos da estrutura de ensaio proposta neste trabalho, composta por duas ferramentas computacionais, uma de simulação de eventos discretos e uma outra de MP.

A ferramenta de simulação de eventos discretos é responsável por gerar os registros contendo todos os eventos ocorridos quando se executa uma simulação de um processo previamente modelado. A ferramenta de simulação de eventos discretos escolhida neste trabalho foi o *Arena Simulation Software*⁴ por permitir configurações que possibilitem registrar e exportar dados da simulação. Por outro lado, a ferramenta de MP é utilizada para extrair conhecimento do RE criado previamente com a simulação do modelo de processo escolhido. Observe que a única informação do modelo simulado que é acessada pela MP é o RE. Para a MP foi escolhido o *Process Mining Framework 6.9* (ProM) ⁵, devido ao programa permitir diversas técnicas de MP e visualizações dos resultados, além de ser um programa gratuito.

As parametrizações e os detalhes das configurações realizadas no *Arena Simulation*Software para a geração do RE artificial são apresentadas a seguir.

٠

⁴ https://www.arenasimulation.com

⁵ http://www.processmining.org



Figura 8 – Estrutura de ensaio proposta Fonte: Autora.

3.1 METODOLOGIA PARA GERAÇÃO DO REGISTRO DE EVENTOS ARTIFICIAL

Para a geração do RE artificial de um processo previamente modelado são necessárias as seguintes etapas: i) conversão de modelos, ii) criação do RE, iii) parametrização dos módulos, iv) geração do arquivo externo e, finalmente, v) formatação do arquivo externo. As etapas de i a iv são realizadas pelo Arena e a etapa v pelo *Microsoft Office Excel* pela facilidade de edição e formatação dos dados em formato .csv.

3.1.1 Conversão de modelos

A etapa de conversão de modelos é responsável pela criação do modelo de simulação interpretável pelo Arena a partir do modelo de processo BPMN previamente desenhado. Essa etapa é realizada de forma manual através da criação na área de trabalho do Arena do modelo de simulação contendo os módulos e interligações equivalentes ao modelo BPMN. Na Tabela 3 estão representadas as equivalências entre os módulos do modelo BPMN e do modelo de simulação usado no Arena.

Módulo	Função	BPMN
Create	Permite ao programa a criação de entidades que serão processadas na simulação	Start event
Process	Principal módulo da simulação, representa o recurso que realiza o processamento das entidades	Task 1

Oecide	Módulo responsável por tomadas de decisões no processo, decisões podem ser baseadas em atributos das entidades, ou	Gateway
	probabilidades. Módulo que representa a chegada da entidade ao ponto final.	End event
Dispose	modulo que representa a enegada ea entidade do ponto inida.	O
Assign	Módulo usado para atribuir novos valores às entidades, podendo ser atributos, variáveis, tipos, imagens, etc.	Não disponível no BPMN
ReadWrite	Graças a esse módulo é possível ler informações de atributos e variáveis da simulação e gravá-las em um arquivo externo.	Não disponível no BPMN
File	Módulo para acessar arquivos externos para usar com o ReadWrite. Com o módulo é possível definir os tipos de acesso e características operacionais.	Não disponível no BPMN

Tabela 3 – Comparação dos módulos do Arena com o BPMN Fonte: Autora

A ilustração de uma conversão de um modelo BPMN para um modelo de simulação usando a equivalência indicada na Tabela 3, pode ser visto na Figura 9. Note que os blocos que representam no modelo BPMN *star event*, *task*, *gateway* e *end event* foram substituídos no modelo de simulação pelos blocos *create*, *process*, *decide* e *dispose*, respectivamente.

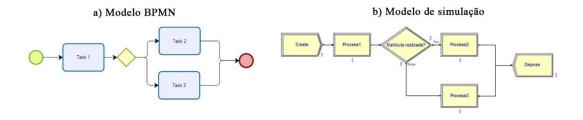


Figura 9 – Exemplo da conversão de modelos BPMN-Arena Fonte: Autora.

As próximas etapas, compostas por ii) criação do RE, iii) parametrização dos módulos e iv) geração arquivo externo são responsáveis pela criação de um arquivo externo contendo todo o RE. A formatação do arquivo externo utilizada neste trabalho contendo parte do RE simulado pode ser observada na Tabela 4. Os campos de dados que compõem o RE apresentado na Tabela 4 são suficientes para os propósitos deste trabalho, ou seja, realizar o processo de descobrimento através da MP aplicado a um RE gerado artificialmente.

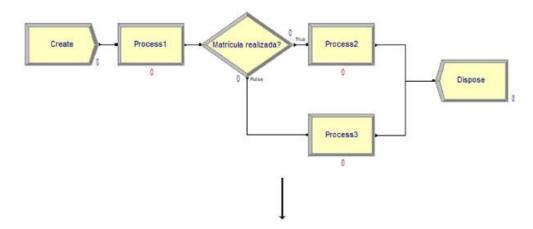
			Event	
Tf	ProcessNumber	Case	ID	Resource
00/01/1900 01:36:43	1	3	1	1111
00/01/1900 01:50:03	1	4	2	1111
00/01/1900 01:52:41	1	2	3	1111
00/01/1900 02:07:47	3	4	4	3333
00/01/1900 02:17:08	3	2	5	3333
00/01/1900 02:20:29	1	5	6	1111
00/01/1900 03:31:56	2	3	7	1111
00/01/1900 03:47:41	1	6	8	1111
00/01/1900 03:54:14	1	7	9	1111
00/01/1900 04:15:06	3	6	10	3333
00/01/1900 04:33:33	2	5	11	1111
00/01/1900 04:46:31	1	8	12	1111
00/01/1900 05:28:19	2	7	13	1111
00/01/1900 05:31:36	1	9	14	1111
00/01/1900 05:54:30	3	9	15	3333
00/01/1900 06:06:58	2	8	16	1111
00/01/1900 07:41:27	1	10	17	1111
00/01/1900 07:44:51	1	12	18	1111
00/01/1900 08:04:50	3	10	19	3333
00/01/1900 08:26:00	1	11	20	1111
00/01/1900 08:52:36	2	12	21	1111

Tabela 4 – RE simulado a partir do Arena Fonte: Autora.

3.1.2 Criação do registro de eventos

Nessa etapa são incluídos os módulos que permitem ao Arena registrar os dados da simulação em um arquivo externo. Os módulos responsáveis por essa tarefa são os módulos *Assign* e o *ReadWrite* apresentados na Tabela 3. Um exemplo do resultado da adição desses dois módulos em um modelo de simulação do Arena pode ser visto na Figura 10. Observe que o conjunto formado pelos módulos *Assign* e *ReadWrite* são inseridos, respectivamente, a jusante de cada módulo *Process* do modelo de simulação do Arena discutido na seção 3.1.1.

Modelo de simulação



Modelo de simulação para registro de eventos

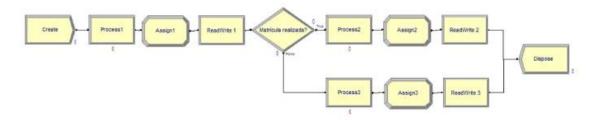


Figura 10 – Adição de módulos para RE Fonte: Autora

3.1.3 Parametrização dos módulos

Alguns módulos usados no modelo de simulação do Arena precisam ser parametrizados para simular de forma satisfatória o modelo de processos representado pelo modelo BPMN inicial. Em outros módulos, a parametrização é fixa e independe do modelo BPMN em estudo. Além destes, tem-se também o módulo *Dispose* que não requer qualquer parametrização.

A parametrização dos módulos *Create*, *Process* e *Decide*, dependem dos atributos de seus equivalentes no modelo BPMN. As parametrizações necessárias destes módulos são:

- 1) Módulo *Create*: configuração da taxa de criação das entidades, que são as instâncias que irão percorrer o processo simulado;
- 2) Módulo *Process*: configuração da distribuição de probabilidade da variável aleatória que representa o tempo de execução daquela tarefa;
- 3) Módulo Decide: configuração da condição ou probabilidade de desvio do fluxo.

Por outro lado, a parametrização dos módulos *Assign*, *ReadWrite* e *File* são independentes do modelo BPMN que se pretende analisar. Com isso, uma vez escolhida a formatação desejada do registro de eventos, as parametrizações destes módulos são então realizadas e não se alteram com a mudança do modelo BPMN em estudo.

3.1.3.1 Parametrização do registro de eventos

Como indicado na Tabela 3, o módulo *Assign* é utilizado para registrar no sistema, a cada iteração da simulação, os dados que identificam a realização da tarefa precedente que serão futuramente gravados no RE, ou seja, o nome da tarefa (*varTaskName*), o recurso (*varResourceName*) que a realizou e um contador para identificar o evento daquela tarefa realizada (*varEventCounter*). Esses dados serão responsáveis pelos valores identificados no RE (Tabela 4) como, respectivamente, "ProcessNumber", "Resource" e "Event ID". A Figura 11 ilustra de forma resumida o relacionamento dos blocos *Assign* e *ReadWrite* para a criação do RE.

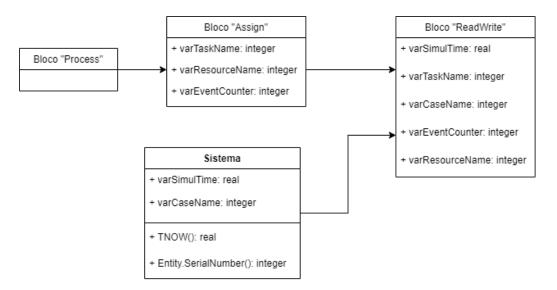


Figura 11 – Diagrama de relacionamento dos blocos *Assign* e *ReadWrite* para a criação do registro de eventos Fonte: Autora

O módulo *ReadWrite* reúne as informações geradas pelo módulo *Assign* (*varTaskName*, *varResouceName* e *varEventCounter*) além de informações da simulação, como o tempo de simulação atual (*varSimulTime*), através da função *TNOW*, e o índice da entidade que gerou o evento (*varCaseName*), através da função *Entity.SerialNumber*. Esses dados são identificados no RE (Tabela 4) como, respectivamente, "Time" e "Case".

Na Tabela 5 estão indicadas as parametrizações necessárias para cada módulo Assign, em que TName, RName, $varEventCounter \in \mathbb{Z}$. TName indica o nome atribuído à tarefa e, RName o nome atribuído ao recurso, ambos referentes ao módulo Process que precede ao módulo que está sendo configurado.

Type	Name	New Value	
Attribute	varTaskName	TName	
Attribute	varResourceName	RName	
Variable	varEventCounter	varEventCounter + 1	

Tabela 5 – Configuração do módulo *Assign* Fonte: Autora.

Na Tabela 6 estão indicadas as parametrizações necessárias para cada módulo ReadWrite. Uma vez que esse módulo é responsável por alocar as informações no arquivo externo, a ordem em que cada parâmetro é inserido no módulo ReadWrite coincide com a ordem das colunas que compõem o RE indicado na Tabela 4, ou seja, a ordem apresentada na Tabela 6.

Type	Variable/Attribute/ Other	Format
Other	TNOW	Real decimal
Attribute	varTaskName	Integer
Other	Entity.SerialNumber	Integer
Variable	varEventCounter	Integer
Attribute	varResourceName	Integer

Tabela 6 – Configuração do módulo *ReadWrite* Fonte: Autora.

É importante atentar-se que as informações das Tabela 5 e Tabela 6, referem-se ao módulo *Process* imediatamente anterior aos módulos *Assign* e *ReadWrite* que estão sendo configurados. O módulo *ReadWrite* uma vez programado, seguirá os mesmos parâmetros para qualquer módulo *Process* adicionado no modelo de processo. A única alteração a ser feita é a identificação de cada módulo. A conexão dos dados coletados pelo módulo *ReadWrite* com um arquivo externo ao programa Arena é feita através do módulo *File* (Tabela 3). Este módulo não aparece no modelo de simulação como os outros módulos discutidos anteriormente, mas é preciso ativá-lo, visto que tem um papel importante na geração do RE: o módulo indica o arquivo onde as informações do módulo *ReadWrite* serão alocadas, bem como o título e a quantidade de colunas que serão utilizadas.

3.1.4 Geração do arquivo externo

O arquivo externo contém o RE completo da simulação realizada. Ele possui a formatação indicada na Tabela 4 e é o elo de ligação entre o modelo BPMN inicialmente utilizado e a ferramenta computacional de MP. Para a obtenção do arquivo externo, denominado neste trabalho de RE artificial, inicialmente é criado um arquivo externo cujos dados são separados por vírgulas, tipo CSV (*comma-separated value*), no computador em que executa a simulação, e então, usa-se o módulo *File* para conectar o diretório do arquivo CSV criado com o programa *Arena Simulation*.

O módulo *File* admite outros tipos de arquivos externos, além do CSV. Neste trabalho optou-se pelo tipo CSV pela facilidade de utilização dessa extensão no programa de MP que será utilizado. Este módulo é responsável também pela criação e nomeação das colunas que serão usadas no arquivo CSV. Na Tabela 7, a coluna *Assignments ReadWrite* corresponde a cada atribuição do módulo *ReadWrite* (Tabela 6) e a coluna *Column Header Text* corresponde à parametrização do módulo *File* com os nome escolhidos para o cabeçalho do RE.

Assignments	Column		
ReadWrite	Header Text		
1	Tf		
2	ProcessName		
3	Case		
4	EventID		
5	Resource		

Tabela 7 – Parametrização do módulo *File*Fonte: Autora.

O resultado das parametrizações dos módulos do Arena que foram discutidas nessa seção após a simulação do processo é mostrado na Tabela 8.

Tf	ProcessNumber	Case	Event ID	Resource
0,067161	1	3	1	1111
0,076426	1	4	2	1111
0,078257	1	2	3	1111
0,088739	3	4	4	3333
0,095235	3	2	5	3333
0,097555	1	5	6	1111
0,14718	2	3	7	1111
0,158112	1	6	8	1111
0,16266	1	7	9	1111

0,177153	3	6	10	3333
0,189966	2	5	11	1111
0,198973	1	8	12	1111
0,227995	2	7	13	1111
0,230274	1	9	14	1111
0,246176	3	9	15	3333
0,254841	2	8	16	1111
0,320454	1	10	17	1111
0,322817	1	12	18	1111
0,336687	3	10	19	3333
0,351391	1	11	20	1111
0,369856	2	12	21	1111

Tabela 8 – RE resultado da simulação no Arena Fonte: Autora.

Observe que na Tabela 8 os dados da coluna "Tf" é do tipo real que é incompatível com a ferramenta de MP utilizada.

3.1.5 Formatação do arquivo externo

O programa *Microsoft Office Excel* é usado para formatar os dados do RE gerado pelo Arena que são incompatíveis com o ProM. O ProM lê diversas formatações de tempo, exceto informações de tempo do tipo real. Uma vez que optou-se nesse trabalho pela formação de tempo do tipo dd/MM/aaaa HH:mm:ss, por ser uma formatação completa que agrega data e hora, tem-se a necessidade de transformar os dados do tipo real da coluna "Tf" do RE para o tipo escolhido. Tal formatação é feita através da função de formatação numérica no *Excel*. A escolha desse programa dá-se uma vez que o programa já é amplamente utilizado e permite a alteração de células bem como a gravação do arquivo em extensão CSV. Com a alteração da coluna "Tf" é obtido, então, o RE artificial usado para fazer a MP como mostrado na Tabela 4.

O RE artificial gerado tem como base os conceitos vistos na seção 2.3, ou seja, as informações contidas no registro são pontuais e individuais de cada atividade executada pela simulação, sem qualquer informação explícita sobre o modelo de processo que foi utilizado como base.

Este capítulo apresentou as etapas para a criação do RE artificial a partir de um modelo de processo de negócio BPMN. A seguir, com o objetivo de validação da proposta deste trabalho de geração de RE artificial, são apresentados os resultados obtidos quando se

submete o RE artificiais gerado pela simulação um modelo de processo previamente escolhido à técnica de descobrimento na MP.

4. RESULTADOS SIMULADOS

Para validar a proposta original deste trabalho, obtenção de um RE artificial de um dado processo de negócio, foi utilizada uma ferramenta de MP utilizando a técnica de descobrimento que foi discutida na Seção 2.5. Esta técnica extrai conhecimento do RE de um processo apresentando como resultado o modelo de processo que gerou aquele RE. Como discutido na Seção 3 a ferramenta de MP utilizada foi o *Process Mining Framework 6.9* (ProM). Dentre os algoritmos de MP utilizados no ProM, foi escolhido o *alpha++* (VAN DER AALST, W. M. P.; WEIJTERS; MARUSTER, 2004) por ser uma melhoria do algoritmo *alpha* (WEN; W. M.P. AALST; *et al.*, 2007), um dos primeiros algoritmos apresentados na literatura da MP, possibilitando descobrir não só pequenos loop e loops próprios, como também escolhas com sincronização como acontece na maioria dos processos na vida real.

4.1 MODELO SIMULADO

O modelo de processo de negócio que foi utilizado neste capítulo para validar o RE artificial proposto pode ser observado na Figura 12. Este modelo de processo é do tipo *lasagna*.

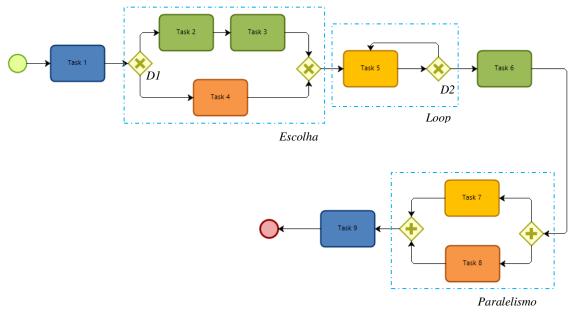


Figura 12 – Processo utilizado para testar o RE artificial

O processo escolhido contém elementos importantes e comuns para modelagem de processos como: escolha, loop e paralelismo. Tais restrições precisam ser captadas pelo RE e pelo algoritmo de MP utilizado. A inclusão desses elementos de interação entre os processos garante que boa parte de modelos de processos tipo *lasagna* serão contemplados com a proposta deste trabalho de geração do RE artificial.

No processo proposto, existe um ramo principal que pode variar seus traces de acordo com as escolhas ou ordem em que são realizadas as atividades. Atividades em paralelo significam que ambas serão executadas, mas a ordem de execução não é pré-determinada. As diferentes cores entre os símbolos que representam os processos explicitam a divisão dos recursos. Quatro recursos são usados no total: o primeiro executa "Task 1" e "Task 9"; o segundo executa os processos "Task 2", "Task 3" e "Task 6"; o terceiro executa "Task 4" e "Task 8"; e o último recurso executa, finalmente, "Task 5" e "Task 7". Considerando que as tarefas competem pelos recursos, não é possível que as atividades que utilizam o mesmo recurso sejam executadas simultaneamente. A escolha e divisão dos recursos foi feita de forma arbitrária, buscando diversidade para que o modelo seja representativo e contemple características gerais de modelo de processos.

4.2 PARÂMETROS DOS MODELOS

O modelo criado também atende uma divisão de probabilidade estatística de execução de acordo com a Tabela 9 e tem uma taxa de tempo entre chegadas distribuída exponencialmente com parâmetro $\lambda = 7,5$ entidades/minuto.

Nome do	Tipo de	Mínimo	Média	Máximo	Recurso
símbolo	distribuição	(min)	(min)	(min)	
Task 1	Triangular	0,5	1	1,5	1
Task 2	Triangular	1	2	2,5	2
Task 3	Triangular	0,5	1	1,5	2
Task 4	Triangular	2	3	4	4
Task 5	Triangular	0,5	1	2	3
Task 6	Triangular	0,7	1	1,5	2
Task 7	Triangular	2	3	4	3
Task 8	Triangular	0,5	2	3	4
Task 9	Triangular	1	3	4	1

Tabela 9 – Distribuição de probabilidade dos processos do modelo Fonte: Autora.

As estatísticas foram escolhidas considerando a limitação do *Arena Software*, versão gratuita, que pode operar com, no máximo, 150 entidades simultaneamente no programa. Os valores foram escolhidos arbitrariamente com o intuito de representar um modelo genérico. As decisões do modelo, representadas como D1 e D2, foram configuradas, também de forma arbitrária, para representar a probabilidade do evento "verdadeiro" como sendo, respectivamente, 65% e 75%.

4.3 PARÂMETROS DA SIMULAÇÃO

Os processos foram simulados considerando uma única replicação com um tempo de simulação equivalente a 100 horas e um período de aquecimento de 2 minutos, gerando, exatamente, 6.227 eventos e 812 casos. Parte do registro de eventos obtido quando o processo é simulado com os parâmetros apresentados neste capítulo pode ser visto na Figura 14.

. Tf	ProcessName	Case	EventID	ResourceName
01/01/1900 00:22:10	1	2	1	1111
01/01/1900 01:18:22	4	2	2	3333
01/01/1900 01:20:07	1	3	3	1111
01/01/1900 01:56:52	1	4	4	1111
01/01/1900 01:59:19	5	2	5	3333
01/01/1900 03:28:08	4	3	6	3333
01/01/1900 04:27:36	4	4	7	3333
01/01/1900 04:54:09	5	2	8	3333
01/01/1900 05:17:29	5	3	9	3333
01/01/1900 05:38:19	5	4	10	3333
01/01/1900 06:08:54	6	2	11	3333
01/01/1900 06:44:07	6	3	12	3333
01/01/1900 07:20:24	5	4	13	3333
01/01/1900 07:33:33	7	2	14	5555
01/01/1900 07:40:34	1	5	15	1111
01/01/1900 08:03:57	8	2	16	4444
		•••	•••	
09/04/1900 23:40:29	2	812	6227	2222

Tabela 10 – Parte do registro de eventos gerado pela simulação após tratamento da coluna Tf Fonte: Autora.

4.4 PARÂMETROS DO PROM

Após a formatação da coluna "Tf", como explicitado na seção 3.1.5, o RE foi filtrado. A etapa de filtragem é necessária para impedir que ruídos e casos incompletos poluam o modelo final. A filtragem, que utiliza o filtro *Simple Heuristics*, elimina do RE os traços que não possuem início e o fim coincidentes com os dos processos simulados, evitando a poluição do modelo que será descoberto. A aplicação do filtro resultou em 6.188 eventos e 805 casos que representa uma diminuição de, aproximadamente, 0,86% dos casos e 0,63% dos eventos contidos no RE. Os baixos percentuais de dados eliminados do RE não interfere significativamente nas análises realizadas.

4.5 RESULTADOS MINERAÇÃO DE PROCESSOS

Na Figura 13 pode-se observar as quantidades de ocorrência por cada atividade do modelo proposto, obtidas através do ProM logo após a filtragem.

Número de ocorrência por processo 1080 805 805 805 805 805 278 278 278 Task 1 Task 2 Task 3 Task 4 Task 5 Task 6 Task 7 Task 8 Task 9

Figura 13 – Número de ocorrência por processo Fonte: Autora.

Inspecionando o modelo proposto, pode-se ver que os processos em série "Task 2" e "Task 3" são concorrentes com o "Task 4", o que vai ao encontro dos números indicados na Figura 13, em que dos 805 (100%) casos simulados, 278 (34,5%) foram processados no ramo contendo "Task 2" e "Task 3" e 527 (65,5%) casos processados no "Task 4". Observe que o resultado percentual obtido devido à concorrência também concorda com o percentual de decisão inicial atribuído ao D1, a menos de 0,5% que pode ser atribuído ao conjunto finito de casos gerados e ao erro inserido com o processo de filtragem.

Quanto à D2, que determina o loop em "Task 5", apesar da quantidade de casos simulados ser igual a 805, esse processo aconteceu 1.080 vezes, um aumento de 34,16% na ocorrência, aproximadamente, 9% acima da probabilidade estipulada inicialmente para D2, que era de 25%. Essa diferença pode ser atribuída à possibilidade de haver mais de uma recorrência no loop em alguns casos.

Pode-se verificar também na Figura 13, que há uma concordância dos números com o modelo proposto em relação a quantidade de eventos ocorridos nos processos em série "Task 6", "Task 9" e os processos em paralelos "Task 7" e "Task 8".

De uma forma geral, pode-se admitir que os resultados indicados na Figura 13 são compatíveis com o comportamento modelado pelo processo de negócio proposto, ou seja, os resultados sugerem que o RE artificial gerado é compatível com o modelo de processo proposto.

O modelo de processos obtido pela técnica de descobrimento através do algoritmo alpha++ usando a MP está indicado na Figura 14, representado pela notação de Rede de Petri.

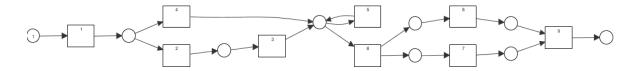


Figura 14 – Resultado da MP usando o algoritmo *alpha*++ Fonte: Autora.

Pode-se observar que, considerando apenas a representação gráfica deste resultado (Figura 14) com a do modelo inicial proposto na Figura 12, é a existência de um loop no processo 5 da Rede de Petri. Esta representação sugere que pode ocorrer alguma situação em que a atividade 6 é processada sem que antes seja processada a atividade 5. Porém, esta possibilidade não existe no modelo BPMN (Figura 12) usado para gerar o RE.

Esse comportamento observado, que desviou do processo proposto, é recorrente nos processos que são descobertos usando o *alpha*++, isso porque no algoritmo, todos os loops encontrados são diretamente conectados a um lugar e não a uma tarefa (VAN DER AALST, W. M. P.; WEIJTERS; MARUSTER, 2004).

5. CONCLUSÕES

Este trabalho apresentou os conceitos de uma ferramenta que surgiu juntamente com a indústria 4.0 e que vem sendo facilitada pela grande disponibilidade de dados e informações nos dispositivos e softwares: a MP. A MP é uma ferramenta que faz ponte entre a mineração de dados, já que ela processa e ordena uma grande quantidade de dados que, inicialmente, parecem desestruturados e a análise de modelos de processos, uma vez que a mineração tem como um dos possíveis *outputs*, o modelo de processos. Com essa ferramenta é possível descobrir o modelo de processos a partir de um RE (técnica de descobrimento), é possível comparar o modelo planejado com o modelo que é executado (técnica de conformidade) e é possível melhorar o processo a partir das análises de performance e conformidade (aprimoramento).

O RE é uma importante entrada da MP e pode ser definido como um registro que contém informações sobre a execução de cada evento em um processo e cada um desses eventos armazena informações quanto ao caso à que ele refere, à atividade que foi executada e um horário ou data de execução. Apesar da indústria 4.0 ter potencializado a criação e armazenamento de dados, nem sempre um RE é fácil de ser obtido, seja por falta da possibilidade do armazenamento desses dados gerados pelo software ou pela máquina, seja o fato do armazenamento do RE não ter todas as características necessárias para a MP, ou pela necessidade de inúmeras autorizações para compartilhamento e uso dos dados.

Entendendo essa dificuldade, esse trabalho propõe uma forma de gerar, através de sua simulação, um registro de evento de um modelo de processo previamente arquitetado com a finalidade de estudar as técnicas de MP.

Primeiramente, foi construído um modelo de processos usado como base para montar a simulação no programa *Arena Simulation*, configurando os atributos desejados como tempo final, nome da atividade, identificação da atividade, número do caso e o recurso que realizou o processo. A simulação gerou um RE com 6.227 eventos e 812 casos. Após a simulação, os dados relativos ao tempo foram formatados para otimizar a MP no programa *Microsoft Excel*. Com o RE configurado, foi possível filtrá-lo, resultando em 6.188 eventos e 805 casos, evitando processos incompletos.

Os resultados apresentados obtidos neste trabalho sugerem que existe uma alta semelhança entre o modelo que foi projetado para a simulação e o que foi gerado pela MP,

ou seja, o RE simulado é capaz de exprimir de forma clara a simulação projetada, podendo ser usado para auxiliar no estudo dos algoritmos de MP, uma vez que a única variação observada no resultado do modelo da MP foi em uma atividade de *loop* do processo, que é derivada de uma limitação da notação da Rede de Petri do algoritmo *alpha*++.

Dada a dificuldade de conseguir um RE, a metodologia apresentada neste trabalho para gerar um RE artificiais pode ser usada gerar um dos principais insumos para estudos e desenvolvimentos de novos algoritmos e outras análises dentro da MP, como estudos de melhoria do modelo, análise de gargalos, e outros objetivos que a MP vem apresentando.

6. REFERÊNCIAS

ABEPRO - Associação Brasileira de Engenharia de Produção | A Profissão. Disponível em: http://portal.abepro.org.br/a-profissao/>. Acesso em: 7 nov. 2019.

AGUIRRE, Jimmy Armas *et al.* Evaluation of operational process variables in healthcare using process mining and data visualization techniques. In: LACCEI INTERNATIONAL MULTI-CONFERENCE FOR ENGINEERING, EDUCATION AND TECHNOLOGY, jul. 2019, [S.l: s.n.], jul. 2019.

BAKRI, Anas A. Al; STEEL, Aileen Cater; SOAR, Jeffery. The influence of B2B e-commerce on SMEs' performance and efficiency: a review of the literature. **International Journal of Liability and Scientific Enquiry** v. 3, n. 3, p. 213, 2010. Disponível em: http://www.inderscience.com/link.php?id=33356>. Acesso em: 13 out. 2020.

BALDAM, Roquemar; VALLE, Rogério; ROZENFELD, Henrique. **Gerenciamento de processos de negócio - BPM: uma referência para implantação prática**. [S.l.]: GEN LTC, 2014. 403 p. .

DAVENPORT, Thomas H. **Reengenharia de Processos**. [S.l.]: Elsevier, 1994. 408 p. .85-7001-874-6.

DE MEDEIROS, A. K. A. *et al.* Process Mining for Ubiquitous Mobile Systems: An Overview and a Concrete Algorithm. In: BARESI, Luciano *et al.* (Orgs.). **Ubiquitous Mobile Information and Collaboration Systems**. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. 3272 v. p. 151–165. Disponível em:

http://link.springer.com/10.1007/978-3-540-30188-2_12. Acesso em: 20 out. 2020. 978-3-540-24100-3.

DUMAS, Marlon; VAN DER AALST, Wil M. P.; TER HOFSTEDE, Arthur H. M. (Orgs.). **Process-Aware Information Systems: Bridging People and Software through Process Technology**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2005. Disponível em: http://doi.wiley.com/10.1002/0471741442. Acesso em: 2 out. 2020. .978-0-471-74144-2.

EFFENDI, Yutika Amelia; SARNO, Riyanarto. PARALLEL PROCESS DISCOVERY USING A NEW TIME-BASED ALPHA++ MINER. **IIUM Engineering Journal** v. 21, n. 1, p. 126–141 , 20 jan. 2020. Disponível em: https://journals.iium.edu.my/ejournal/index.php/iiumej/article/view/1173. Acesso em: 20 out. 2020.

FRANCISCO, Rosemary; SANTOS, Eduardo A Portela. Aplicação da Mineração de Processos como uma prática para a Gestão do Conhecimento. In: SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO, 2011, Salvador. **Anais**... Salvador: [s.n.], 2011. p.9.

GANDULFO, Pablo Ignacio. Método de Mineração de Processos para Auxílio à Tomada de Decisão: um Estudo de Caso no Controle de Férias. p. 113, 2016.

GÖLZER, Philipp; CATO, Patrick; AMBERG, Michael. Data processing requirements of Industry 4.0 - use cases for Big Data applications. 1 jan. 2015, [S.l: s.n.], 1 jan. 2015.

HARRINGTON, H. James; ESSELING, K. C.; NIMWEGEN, Van. Business Process Improvement Workbook: Documentation, Analysis, Design, and Management of Business Process Improvemen. [S.l.]: McGraw-Hill Professional, 1997.

IVERS, Anne Marie; BYRNE, James; BYRNE, Pj. Analysis of SME data readiness: a simulation perspective. **Journal of Small Business and Enterprise Development** v. 23, n. 1, p. 163–188 , 15 fev. 2016. Disponível em: https://www.emerald.com/insight/content/doi/10.1108/JSBED-03-2014-0046/full/html. Acesso em: 5 out. 2020.

JICHA, Arthur; PATTON, Mark; CHEN, Hsinchun. SCADA honeypots: An in-depth analysis of Conpot. In: 2016 IEEE CONFERENCE ON INTELLIGENCE AND SECURITY INFORMATICS (ISI), set. 2016, Tucson, AZ, USA. Anais... Tucson, AZ, USA: IEEE, set. 2016. p.196–198. Disponível em: http://ieeexplore.ieee.org/document/7745468/. Acesso em: 13 out. 2020. 978-1-5090-3865-7. .

JUNIOR, Gilmar Teixeira. Modelagem de Sistemas de Informação para Mineração de Processos: Características e Propriedades das Linguagens. p. 129, 2017.

LOPES, Nathália Conceição de Souza. **Modelo de gestão por processos baseado em mineração**. Tese de Mestrado — Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017. 100 p.

NETTO, Paulo Oswaldo Boaventura; JURKIEWICZ, Samuel. **Grafos: introdução e prática**. [S.l.]: Blucher, 2009. 176 p. . Acesso em: 25 nov. 2020. .85-212-1133-3.

NGAI, E.W.T.; XIU, Li; CHAU, D.C.K. Application of data mining techniques in customer relationship management: A literature review and classification. **Expert Systems with Applications** v. 36, n. 2, p. 2592–2602, mar. 2009. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0957417408001243. Acesso em: 13 out. 2020.

PAMPONET, Arnaud P. *Como entender os processos organizacionais* .administradores.com. [S.l: s.n.]. Disponível em: https://administradores.com.br/artigos/como-entender-os-processos-organizacionais, set. 2019

R ADDO-TENKORANG; P HELO. Enterprise Resource Planning (ERP): A Review Literature Report., 2011. Disponível em: http://rgdoi.net/10.13140/2.1.3254.7844. Acesso em: 13 out. 2020.

SALIMIFARD, Khodakaram; WRIGHT, Mike. Petri net-based modelling of workflow systems: An overview. **European Journal of Operational Research** v. 134, n. 3, p. 664–676 , 1 nov. 2001. Disponível em: http://www.sciencedirect.com/science/article/pii/S0377221700002927>. Acesso em: 18 jun. 2019.

SMART, P. A.; MADDERN, Harry; ROGGER, Maull. Understanding Business Process Management: Implications for Theory and Practice. **British Journal of Management** p. 491–507, dez. 2009.

SONG, Minseok; GÜNTHER, Christian W.; VAN DER AALST, Wil M. P. Trace Clustering in Process Mining. In: ARDAGNA, Danilo; MECELLA, Massimo; YANG, Jian (Orgs.). . **Business Process Management Workshops**. Lecture Notes in Business Information Processing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. 17 v. p. 109–120.

Disponível em: http://link.springer.com/10.1007/978-3-642-00328-8_11. Acesso em: 13 out. 2020. 978-3-642-00327-1.

SURIADI, S. *et al.* Event log imperfection patterns for process mining towards a systematic approach to cleaning event logs. In: INFORMATION SYSTEMS - DATABASES: CREATION, MANAGEMENT AND UTILIZATION, jul. 2016, [S.l.]: Elsevier, jul. 2016. p.132–150. Disponível em: https://doi.org/10.1016/j.is.2016.07.011.

UNERTL, Kim M *et al.* Traversing the many paths of workflow research: developing a conceptual framework of workflow terminology through a systematic literature review. **Journal of the American Medical Informatics Association** v. 17, n. 3, p. 265–273, maio 2010. Disponível em: https://academic.oup.com/jamia/article-lookup/doi/10.1136/jamia.2010.004333. Acesso em: 13 out. 2020.

VAN DER AALST, W. M. P. *et al.* Business process mining: An industrial application. **Information Systems** v. 32, n. 5, p. 713–732 , 1 jul. 2007. Disponível em: http://www.sciencedirect.com/science/article/pii/S0306437906000305. Acesso em: 7 jun. 2019.

VAN DER AALST, W. M. P.; WEIJTERS, Ton; MARUSTER, Laura. Workflow Mining: Discovering Process Models from Event Logs. **IEEE Transactions on Knowledge and Data Engineering** v. 16, n. 9, p. 1128–1142 , set. 2004. Disponível em: https://ieeexplore.ieee.org/document/1316839>. Acesso em: 22 set. 2020.

VAN DER AALST, Wil M. P. **Process Mining: Data Science in Action**. 2. ed. [S.l.]: Springer, 2016. 488 p. .978-3-662-49850-7.

VAN DER AALST, Wil M. P.; VAN HEE, Kees Max. Workflow Management: Models, Methods, and Systems. [S.l.]: MIT Press, 2002. 388 p. .978-0-262-72046-5.

VAN DER AALST, Wil M.P. Process Mining: Discovery, Conformance and Enhancement of Business Processes. 11. ed. [S.l.]: Springer, 2011. .

VAN DER AALST, W.M.P.; WEIJTERS, A.J.M.M. Process mining: a research agenda. **Computers in Industry** v. 53, n. 3, p. 231–244, abr. 2004. Disponível em: https://linkinghub.elsevier.com/retrieve/pii/S0166361503001945>. Acesso em: 5 out. 2020.

WEN, Lijie; W. M.P. AALST, Van Der; *et al.* Mining process models with non-free-choice constructs. **Data Mining and Knowledge Discovery** v. 15, n. 2, p. 145–180, 2007. Disponível em: https://research.tue.nl/en/publications/mining-process-models-with-non-free-choice-constructs. Acesso em: 22 set. 2020.

WEN, Lijie; VAN DER AALST, Wil M. P.; *et al.* Mining process models with non-free-choice constructs. **Data Mining and Knowledge Discovery** v. 15, n. 2, p. 145–180, out. 2007. Disponível em: http://link.springer.com/10.1007/s10618-007-0065-y. Acesso em: 20 out. 2020.

WESKE, Mathias. **Business Process Management: Concepts, Languages, Architectures**. 2. ed. [S.l.]: Springer, 2012. 404 p. .978-3-642-28615-5.

WESTERGAARD, Michael. Process Mining Manifesto. **Business Process Management Workshops 2011** v. 99, n. Springer-Verlag, p. 15, 2011.



Termo de Declaração de Autenticidade de Autoria

Declaro, sob as penas da lei e para os devidos fins, junto à Universidade Federal de Juiz de Fora, que meu Trabalho de Conclusão de Curso do Curso de Graduação em Engenharia de Produção é original, de minha única e exclusiva autoria. E não se trata de cópia integral ou parcial de textos e trabalhos de autoria de outrem, seja em formato de papel, eletrônico, digital, áudio-visual ou qualquer outro meio.

Declaro ainda ter total conhecimento e compreensão do que é considerado plágio, não apenas a cópia integral do trabalho, mas também de parte dele, inclusive de artigos e/ou parágrafos, sem citação do autor ou de sua fonte.

Declaro, por fim, ter total conhecimento e compreensão das punições decorrentes da prática de plágio, através das sanções civis previstas na lei do direito autoral¹ e criminais previstas no Código Penal², além das cominações administrativas e acadêmicas que poderão resultar em reprovação no Trabalho de Conclusão de Curso.

Juiz de Fora, 12 de NOVEMBRO de 2020.

AURA MATOS VASCONCELLOS 201349080

NOME LEGÍVEL DO ALUNO (A) Matrícula

Laura Matos Vasconcellos 123.538.616-36
ASSINATURA CPF

¹ LEI N° 9.610, DE 19 DE FEVEREIRO DE 1998. Altera, atualiza e consolida a legislação sobre direitos autorais e dá outras providências.

² Art. 184. Violar direitos de autor e os que lhe são conexos: Pena - detenção, de 3 (três) meses a 1 (um) ano, ou multa.