

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA

CAIO APPELT REZENDE

Arcabouço de Classificação e Escolha de Algoritmos de Descoberta de Processos

Goiânia
2017

**TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS
DE TESES E
DISSERTAÇÕES NA BIBLIOTECA DIGITAL DA UFG**

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio da Biblioteca Digital de Teses e Dissertações (BDTD/UFG), regulamentada pela Resolução CEPEC nº 832/2007, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: ☒ **Dissertação** ☐ **Tese**

2. Identificação da Tese ou Dissertação:

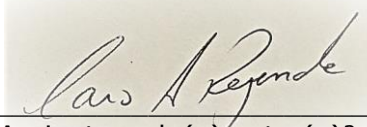
Nome completo do autor: Caio Appelt Rezende

Título do trabalho: Arcabouço de Classificação e Escolha de Algoritmos de Descoberta de Processos

3. Informações de acesso ao documento:

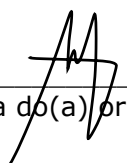
Concorda com a liberação total do documento ☒ **SIM** ☐ **NÃO**¹

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF da tese ou dissertação.



Assinatura do(a) autor(a)²

Ciente e de acordo:



Assinatura do(a) orientador(a)²

Data: 21 / Julho / 2017

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Os dados do documento não serão disponibilizados durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente
- Submissão de artigo em revista científica
- Publicação como capítulo de livro
- Publicação da dissertação/tese em livro

²A assinatura deve ser escaneada.

CAIO APPELT REZENDE

Arcabouço de Classificação e Escolha de Algoritmos de Descoberta de Processos

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Informática da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Programa de Pós-Graduação Stricto Sensu em Ciência da Computação.

Área de concentração: Ciência da Computação.

Orientador: Prof. Dr. Fernando Marques Federson

Goiânia
2017

Ficha de identificação da obra elaborada pelo autor, através do
Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Appelt Rezende, Caio

Arcabouço de Classificação e Escolha de Algoritmos de Descoberta
de Processos [manuscrito] / Caio Appelt Rezende. - 2017.
lxxi, 71 f.: il.

Orientador: Prof. Dr. Fernando Marques Federson.

Dissertação (Mestrado) - Universidade Federal de Goiás, Instituto
de Informática (INF), Ciência da Computação, Goiânia, 2017.
Bibliografia.

Inclui gráfico, tabelas, algoritmos, lista de figuras, lista de tabelas.

1. Processos de negócios. 2. Classificação de algoritmos. 3.
Descoberta de Processos. 4. Sistemas de Informação Gerenciais. 5.
Mineração de Processos. I. Marques Federson, Fernando, orient. II.
Título.

CDU 004



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO



ATA Nº 11/2017

ATA DA SESSÃO DE JULGAMENTO DA DISSERTAÇÃO
DE MESTRADO DE CAIO APPELT REZENDE

Aos três dias do mês de maio de dois mil e dezessete, às dez horas, na sala 151 do Instituto de Informática da Universidade Federal de Goiás, Campus Samambaia, reuniu-se a banca examinadora designada na forma regimental pela Coordenação do Curso para julgar a dissertação de mestrado intitulada “**Arcabouço de Classificação e Escolha de Algoritmos de Descoberta de Processos**”, apresentada pelo aluno Caio Appelt Rezende como parte dos requisitos necessários à obtenção do grau de Mestre em Ciência da Computação, área de concentração Ciência da Computação. A banca examinadora foi presidida pelo orientador do trabalho de dissertação, Professor Doutor Fernando Marques Federson (INF/UFG), tendo como membros os Professores Doutores Anderson da Silva Soares (INF/UFG) e Eliomar Araújo de Lima (INF/UFG). Aberta a sessão, o candidato expôs seu trabalho. Em seguida, o aluno foi arguido pelos membros da banca e:

☒ tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de sua dissertação, a banca concluiu pela **aprovação** do candidato, sem restrições.

☐ não tendo demonstrado suficiência de conhecimento e capacidade de sistematização do tema de sua dissertação, a banca concluiu pela **reprovação** do candidato.

Os trabalhos foram encerrados às 12:00 horas. Nos termos do Regulamento Geral dos Cursos de Pós-Graduação desta Universidade, lavrou-se a presente ata que, lida e julgada conforme, segue assinada pelos membros da banca examinadora.

Prof. Dr. Fernando Marques Federson _____

Prof. Dr. Anderson da Silva Soares _____

Prof. Dr. Eliomar Araújo de Lima _____

Agradecimentos

A Deus, primeiramente, pois, não importa por quão tortos caminhos ande, Ele realiza o impossível em minha vida e me surpreende a cada passo que dou.

A pessoas maravilhosas que insistem em me inspirar e dar motivações, em especial ao meu orientador Prof. Dr. Fernando Marques Federson, que a cada reunião, fazia-me perceber a sede que sinto do saber.

Aos meus pais, Carmen Beatriz Appelt e Renato Miranda Rezende, que me ensinaram o caminho a trilhar e me conduziram por boa parte da estrada.

Aos meus avós, Abílio Appelt e Oswaldo França Rezende, *in memoriam*, por terem sido exemplo de homens honestos, trabalhadores, empreendedores e me motivado a seguir o meu coração nas escolhas que me trouxeram onde cheguei.

Às minhas avós, Elídia Antonia Francescato e Marieta Miranda Rezende, por serem exemplo de mulheres honradas, lutadoras, companheiras e me alimentado com muitos sonhos e vontade de viver.

Aos meus irmãos, Ana Beatriz Appelt Rezende, Luana da Cunha Rezende, Erick da Cunha Rezende, Maria Alice Appelt Brito, que são mais novos que eu e me trazem o sentimento de querer ser um exemplo para eles.

À minha parentela, que, mesmo distante, é motivo de exemplo, ânimo e fé para mim, em especial minha madrinha Adriana Maria Miranda Rezende e tio Max Helmuth Appelt.

A meus chefes e à empresa que trabalho a muitos anos, por apoiarem minha empreitada e a incentivarem, compreendendo que os trabalhos do programa de mestrado várias vezes exigiram minha ausência.

A meus colegas, que muitas vezes me auxiliaram e participaram diretamente no meu desempenho acadêmico, em especial Gilmar Teixeira Júnior, por ter corrido comigo a reta final neste programa de mestrado.

Aos revisores deste trabalho, que além do meu professor e a banca avaliadora, tive o apoio de Afonso Rezende Lopes e Rheder Rosa Landim.

E a todos os envolvidos no processo, desde quem atuou diretamente nele, sejam professores das disciplinas ou secretárias do curso, a quem indiretamente permitiu que este trabalho fosse realizado.

O temor do Senhor é o princípio da ciência.

Salomão,
Provérbios 1:7a.

Resumo

Rezende, Caio Appelt. **Arcabouço de Classificação e Escolha de Algoritmos de Descoberta de Processos**. Goiânia, 2017. 71p. Dissertação de Mestrado. Instituto de Informática, Universidade Federal de Goiás.

A Mineração de Processos (*Process Mining*) é uma área de pesquisa recente e é composta por técnicas que permitem a análise e a extração de conhecimento a partir dos registros de eventos (*logs*) dos processos de negócios obtidos de Sistemas de Informação Gerenciais (*SIG*). As análises podem ser classificadas em três tipos: Descoberta de Processos, Checagem da Conformidade e Melhoria de Processos. Com o atual crescimento não apenas da quantidade, mas também dos tipos de algoritmos que procuram cumprir os objetivos da Mineração de Processos, uma classificação que leve em consideração a performance do algoritmo nas diversas situações reais de sua aplicação se torna importante. A Avaliação e a Comparação dos algoritmos a partir dos dados do repositório poderiam ser feitas através da aplicação de Métricas de Qualidade ou Técnicas de Aprendizado de Máquina. Este trabalho apresenta uma proposta de um conjunto de Métricas de Qualidade que tem como objetivo permitir a classificação, avaliação e comparação de algoritmos de Descoberta de Processos. A proposta foi construída com base na revisão dos algoritmos e suas famílias; no levantamento das possíveis características de performance, que podem ser aplicadas a qualquer tipo de algoritmo sendo testado; e em simulações de registros de eventos de padrões de processos de negócio. Os resultados obtidos pelo trabalho são promissores no sentido de criar a base conceitual e uma metodologia para que futuras pesquisas permitam a construção de um arcabouço (*framework*) de Avaliação e Comparação de novos algoritmos.

Palavras-chave

Processos de negócios, Classificação de algoritmos, Descoberta de Processos, Sistemas de Informação Gerenciais, Mineração de Processos

Abstract

Rezende, Caio Appelt. **Classification and selection of Process Discovery algorithms framework**. Goiânia, 2017. 71p. MSc. Dissertation. Instituto de Informática, Universidade Federal de Goiás.

Process Mining is a recent area of research and is composed of techniques that allow the analysis and extraction of knowledge from the logs of the business processes obtained from Management Information Systems (MIS). The analyzes can be classified into three types: Process Discovery, Conformance Check and Process Improvement. With the current growth not only of quantity, but also of the types of algorithms that seek to fulfill the objectives of Process Mining, a classification that takes into account the performance of the algorithm in the various real situations of its application becomes important. The Evaluation and Comparison of the algorithms from the repository data could be done through the application of Quality Metrics or Machine Learning Techniques. This work presents a proposal of a set of Quality Metrics to allow the classification, evaluation and comparison of Process Discovery algorithms. The proposal is based on the review of algorithms and their families; the possible performance characteristics, that can be applied to any type of algorithm being tested; and in simulations of business process patterns. The results obtained by the work are promising in the sense of creating the conceptual basis and a methodology for future research to allow the construction of a framework for Evaluation and Comparison of new algorithms.

Keywords

Business Process, Algorithms Classification, Process Discovery, Management Information Systems, Process Mining

Sumário

Lista de Figuras	11
Lista de Tabelas	12
1 Introdução	14
1.1 Objetivos	15
2 Fundamentação	16
2.1 Mineração de Processos (<i>Process Mining</i>)	16
2.1.1 Descoberta de Processos (<i>Process Discovery</i>)	18
Redes de Petri	18
Redes de Workflow (<i>WF-net</i>)	18
Definições da Descoberta de Processos	19
Características dos Algoritmos de Descoberta de Processo	20
Critérios de Qualidade de um Modelo de Processo	20
2.1.2 Checagem de Conformidade (<i>Conformance Checking</i>)	21
2.2 Aprendizado de Máquina (<i>Machine Learning</i>)	21
2.2.1 Agrupamentos (<i>Clustering</i>)	22
2.2.2 Seleção de Atributos (<i>Selecting attributes</i>)	22
3 Revisão bibliográfica	23
3.1 Algoritmos de Descoberta de Processo	23
3.1.1 Famílias de Abordagem Computacional	23
3.2 Métricas de Checagem de Conformidade	25
3.3 Modelo de Ambiente de Avaliação e Comparação de Algoritmos	26
3.4 Ambiente Implementado por Wang et al	27
3.5 Ambiente Implementado por Ribeiro et al	28
4 Materiais e Métodos	31
4.1 Metodologia	31
4.2 Ferramentas	33
4.2.1 ProM 6.6	33
Complementos no ProM 6.6 de Descoberta de Processos	33
4.2.2 PLG2	34
4.2.3 CoBeFra	35
4.2.4 WEKA	35

5	Experimentos	37
5.1	Configurações	37
5.1.1	PLG2	37
5.1.2	Algoritmos de Descoberta de Processos	37
5.1.3	Algoritmos de Checagem de Conformidade	38
5.1.4	WEKA	39
5.2	Execução	39
5.2.1	Padrão de Sequência de Atividades	39
5.2.2	Padrão de Paralelismo	44
5.2.3	Padrão de Desvio Exclusivo	49
5.2.4	Padrão de Laço	53
6	Análises	57
6.1	Seleção de Atributos	57
6.2	Padrão de Sequência de Atividades	57
6.3	Padrão de Paralelismo	59
6.4	Padrão de Desvio Exclusivo	60
6.5	Padrão de Laço	61
6.6	Resumo	63
6.7	Agrupamentos	63
6.8	Geral	64
7	Conclusão e Trabalhos Futuros	65
7.1	Conclusão	65
7.2	Trabalhos Futuros	66
	Referências Bibliográficas	67

Lista de Figuras

2.1	Contexto de Mineração de Processos: Descoberta, Checagem de Conformidade e Melhoria de Processos [24].	17
3.1	Ambiente modelo, apresentado por Rozinat et al [23].	26
3.2	Ambiente implementado por Wang et al [36], com acréscimo de comparação de recursos e modelos de redução.	28
3.3	Ambiente implementado por Ribeiro et al [22], com histórico de comparações e avaliações já realizadas para sugestão, parte 1.	29
3.4	Ambiente implementado por Ribeiro et al [22], com histórico de comparações e avaliações já realizadas para sugestão, parte 2.	30
4.1	Ambiente de classificação de algoritmos utilizado.	32
5.1	Desenho do modelo de Processo 1, no padrão de sequência.	39
5.2	Desenho do modelo de Processo 1.1, no padrão de sequência.	41
5.3	Desenho do modelo de Processo 1.2, no padrão de sequência.	42
5.4	Desenho do modelo de Processo 1.3, no padrão de sequência.	43
5.5	Desenho do modelo de Processo 2, no padrão de paralelismo.	44
5.6	Desenho do modelo de Processo 2.1, no padrão de paralelismo.	45
5.7	Desenho do modelo de Processo 2.2, no padrão de paralelismo.	46
5.8	Desenho do modelo de Processo 2.3, no padrão de paralelismo.	47
5.9	Desenho do modelo de Processo 2.4, no padrão de paralelismo.	48
5.10	Desenho do modelo de Processo 3, no padrão de desvio exclusivo.	49
5.11	Desenho do modelo de Processo 3.1, no padrão de desvio exclusivo.	50
5.12	Desenho do modelo de Processo 3.2, no padrão de desvio exclusivo.	51
5.13	Desenho do modelo de Processo 3.3, no padrão de desvio exclusivo.	52
5.14	Desenho do modelo de Processo 4, no padrão de laço.	53
5.15	Desenho do modelo de Processo 4.1, no padrão de laço.	54
5.16	Desenho do modelo de Processo 4.2, no padrão de laço.	55
5.17	Desenho do modelo de Processo 4.3, no padrão de laço.	56

Lista de Tabelas

3.1	Comparativo da apresentação de algoritmos selecionados.	24
3.2	Compilação de métricas e sua relação com os critérios de qualidade [4], [34], [33], [43], [35] e [12].	25
4.1	Algoritmos e complementos utilizados do ProM 6.6 para Descoberta de Processos.	34
5.1	Resultados do experimento no padrão de sequência conforme Figura 5.1.	40
5.2	Resultados do experimento no padrão de sequência conforme Figura 5.2.	41
5.3	Resultados do experimento no padrão de sequência conforme Figura 5.3.	42
5.4	Resultados do experimento no padrão de sequência conforme Figura 5.4.	43
5.5	Resultados do experimento no padrão de paralelismo conforme Figura 5.5.	44
5.6	Resultados do experimento no padrão de paralelismo conforme Figura 5.6.	45
5.7	Resultados do experimento no padrão de paralelismo conforme Figura 5.7.	46
5.8	Resultados do experimento no padrão de paralelismo conforme Figura 5.8.	47
5.9	Resultados do experimento no padrão de paralelismo conforme Figura 5.9.	48
5.10	Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.10.	49
5.11	Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.11.	50
5.12	Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.12.	51
5.13	Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.13.	52
5.14	Resultados do experimento no padrão de laço conforme Figura 5.14.	53
5.15	Resultados do experimento no padrão de laço conforme Figura 5.15.	54
5.16	Resultados do experimento no padrão de laço conforme Figura 5.16.	55
5.17	Resultados do experimento no padrão de laço conforme Figura 5.17.	56
6.1	Métricas relevantes do Experimento 5.1.	58
6.2	Métricas relevantes do Experimento 5.2.	58
6.3	Métricas relevantes do Experimento 5.3.	58
6.4	Métricas relevantes do Experimento 5.4.	58
6.5	Métricas relevantes do Experimento 5.5.	59
6.6	Métricas relevantes do Experimento 5.6.	59
6.7	Métricas relevantes do Experimento 5.7.	59
6.8	Métricas relevantes do Experimento 5.8.	60
6.9	Métricas relevantes do Experimento 5.9.	60
6.10	Métricas relevantes do Experimento 5.10.	60

6.11	Métricas relevantes do Experimento 5.11.	61
6.12	Métricas relevantes do Experimento 5.12.	61
6.13	Métricas relevantes do Experimento 5.13.	61
6.14	Métricas relevantes do Experimento 5.14.	62
6.15	Métricas relevantes do Experimento 5.15.	62
6.16	Métricas relevantes do Experimento 5.16.	62
6.17	Métricas relevantes do Experimento 5.17.	62
6.18	Resumo dos algoritmos que conseguiram equivalência com modelo original por padrão.	63
6.19	Aplicação do componente EM para classificar Algoritmos.	63
6.20	Aplicação do componente EM para classificar Padrão de Fluxo de Trabalho.	64
6.21	Aplicação do componente EM para classificar Família.	64

Introdução

A modelagem de processos de negócio é uma abordagem utilizada para análise e criação de novos processos de maneira estruturada [2]. Suas técnicas são posicionadas como candidatas perfeitas para apoiar os sistemas de informação da era de *big data* [22], em que a informação está disponível em grande quantidade para se extrair dela conhecimento das mais diferentes formas. A Mineração de Processos (*Process Mining*) é uma área de pesquisa recente e é composta por técnicas que atuam a partir dos registros de eventos (*logs*) dos processos de negócios obtidos de Sistemas de Informação Gerenciais (*SIG*) [24]. Os objetivos desse instrumento são descobrir, monitorar e melhorar processos reais ou atuais (como são), ao invés de processos imaginados ou aspirados [4].

A quantidade massiva de informações disponíveis, produzidas pelos SIG, se não forem analisadas, podem se perder, ou perder seu significado para o contexto. Várias ferramentas prometeram soluções para esse mercado que vem se expandindo desde a década de noventa, porém, antes da consolidação de Mineração de Processos, não era possível fazer a conversão e análise de ponta a ponta dentro de um gerenciamento de processos de negócios [24].

A Mineração de Processos é composta de três etapas, *Descoberta de Processos*, *Checagem de Conformidade* e *Melhoria de Processos* [24, 22]. Da primeira etapa, como será possível observar no capítulo 3, uma mesma família de algoritmos, seguindo a separação por abordagem computacional, apresenta resultados distintos. Para os praticantes de Mineração de Processos, apenas a informação da família do algoritmo não permite uma escolha adequada para uma dada situação [22]. Por isso, é importante revisar os algoritmos existentes, observando as evoluções ocorridas e a maneira como foram comparados.

Tentativas de padronização dessa classificação existem na literatura, com contribuições de um ambiente modelo apresentado por Rozinat et al [23] e implementações dele com abordagens de Aprendizado de Máquina, vistos em Wang et al [36] e Ribeiro et al [22].

A razão que justifica uma nova classificação é a divisão atual dos algoritmos em quatro grupos principais com base em sua abordagem computacional [24], que não diz respeito a relevância do algoritmo para possíveis cenários de aplicação e a falta de sucesso

das propostas até então apresentadas, como os próprios autores apontam os motivos.

Nesta pesquisa, analisam-se os algoritmos de Descoberta de Processos utilizando-se da Checagem de Conformidade, por meio dos modelos gerados, para compará-los e classificá-los. A maneira proposta para classificá-los consiste em processos de negócios com padrões de fluxo na entrada dos algoritmos e métricas que avaliam quatro critérios de qualidade de um modelo de processo com base, principalmente, na re-execução dos registros de eventos sobre o modelo. Os critérios avaliados são: aptidão, simplicidade, precisão e generalização.

1.1 Objetivos

Observando a literatura, percebeu-se uma falta de padrão em classificação dos algoritmos, como será destacado no Capítulo 3, principalmente quando novos algoritmos são apresentados. Por isso, o objetivo geral deste trabalho é:

Propôr um ambiente que permita a classificação, avaliação e comparação de algoritmos com métricas relevantes.

E para isso as seguintes perguntas guiaram este trabalho:

1. É possível agrupar os algoritmos ao fixar os padrões de fluxo de trabalho na entrada?
2. Das métricas disponíveis, quais são relevantes e suficientes para agrupar e classificar os algoritmos?
3. Ao utilizar o ambiente proposto, é possível classificar os algoritmos por relevância nos resultados obtidos?

Esta dissertação tem nos próximos capítulos os seguintes tópicos: Fundamentação 2, Revisão Bibliográfica 3, Materiais e Métodos 4, Experimentos 5, Análises 6 e Conclusão e Trabalhos Futuros 7.

Fundamentação

Este capítulo traz conceitos importantes utilizados neste trabalho. Nele, estão os conceitos de Mineração de Processos, com destaque para Descoberta de Processos e Checagem de Conformidade.

2.1 Mineração de Processos (*Process Mining*)

As raízes da Mineração de Processos surgiram na década de cinquenta, com a geração de máquinas de estado finitas a partir de exemplos de sequências. Carl Adam Petri introduziu a primeira linguagem de modelagem que captura adequadamente a concorrência e, logo depois, Mark Gold foi o primeiro a explorar diferentes noções de Aprendizado de Máquina em 1967. Quando Mineração de Dados começou a florescer na década de 90, pouca atenção era despendida a processos. A Mineração de Processos ganhou força e foi sumariamente discutido a partir de um marco em 2003 com a primeira pesquisa (*survey*) na área [24].

Desde então, técnicas de Mineração de Processos foram amadurecendo e apoiadas por várias ferramentas. Além disso, enquanto o foco inicial da disciplina estava apenas na Descoberta de Processos, o espectro se ampliou consideravelmente. Como exemplo, Checagem de Conformidade, Multi perspectivas na Mineração de Processos e Suporte Operacional se tornaram parte integral da principal ferramenta de Mineração de Processos, ProM (apresentado no Capítulo 4.2.1). Tudo isso graças à onipresença dos registros de eventos nos sistemas da informação contemporâneos, permitindo a análise de ponta a ponta na Descoberta do Processo [24].

Mineração de Processos é uma disciplina situada entre Aprendizado de Máquina por um lado e Modelagem e Análise de Processos. Os objetivos da Mineração de Processos são descobrir, monitorar e melhorar processos reais ao extrair conhecimento dos registros de eventos já disponíveis nos sistemas atuais [24].

A Mineração de Processos é composta de três etapas, *Descoberta de Processos*, *Checagem de Conformidade* e *Melhoria de Processos* [24, 22]. Na primeira etapa, os registros de eventos são processados por um algoritmo resultando em um modelo de

processo. Na segunda etapa, são comparados os registros de eventos com um modelo de processos, objetivando encontrar inconformidades e assim validar o modelo para os registros de eventos. Na terceira etapa, existe a evolução do modelo, como apresentação do mesmo em outras perspectivas, ou a sobreposição dos eventos no modelo, para identificação de gargalos ou pontos de fuga, dando base para a criação de alternativas de modelagem que implicam numa melhora nessas perspectivas.

A Figura 2.1, traduzida do livro [24], apresenta esquematicamente a Mineração de Processos, com destaque para a direção em que cada etapa ocorre em relação aos modelos e aos registros de eventos. Note que o "mundo" é representado por várias entidades que interagem com sistemas computacionais, que por sua vez geram os "registros de eventos" nas mais variadas formas. Esses registros nem sempre são estruturados. Com os registros, realiza-se a "descoberta" para se obter o "modelo de processo", numa via de mão única. Reutiliza-se os "registros de eventos" e o "modelo" para a "Checagem de Conformidade", numa via de mão dupla. Por fim, o processo é expandido com "registros de eventos" para gerar novos "modelos", numa via de mão única, porém, retro alimentada.

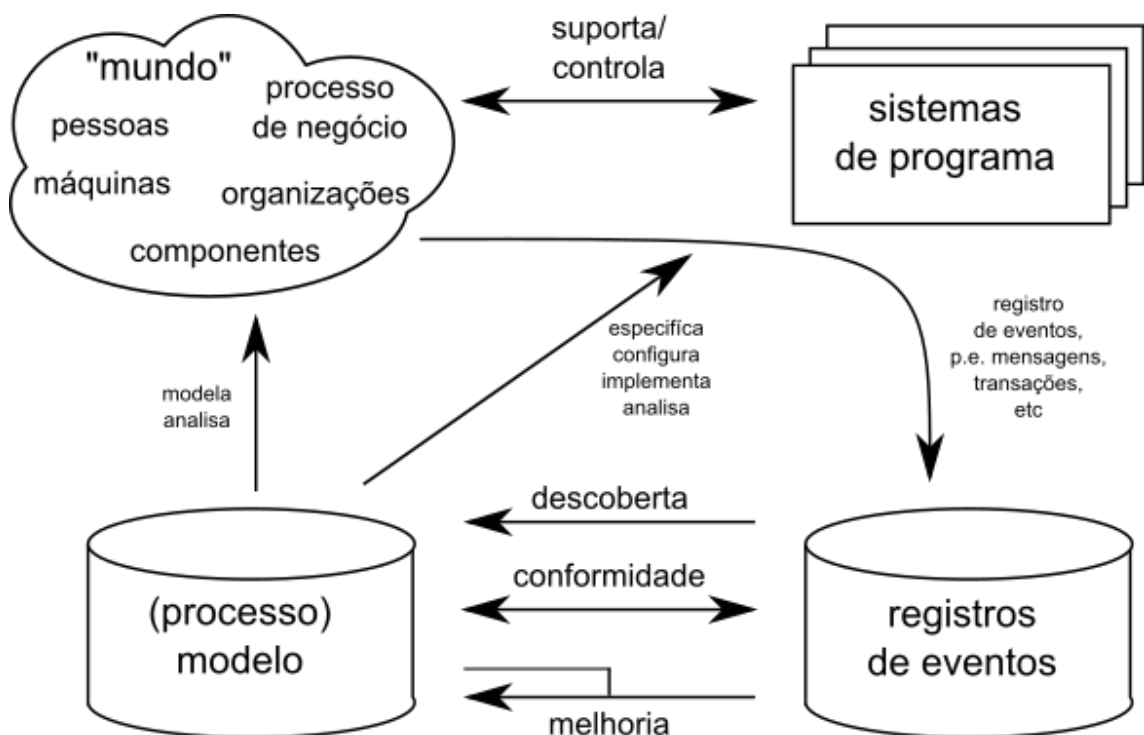


Figura 2.1: Contexto de Mineração de Processos: Descoberta, Checagem de Conformidade e Melhoria de Processos [24].

As duas primeiras etapas serão melhor descritas nas próximas subseções para melhor entendimento do contexto deste trabalho.

2.1.1 Descoberta de Processos (*Process Discovery*)

A Descoberta de Processos é uma das etapas mais desafiadoras da Mineração de Processos. Baseada em registros de eventos, o modelo do processo é construído capturando o comportamento encontrado neles [26].

Seguem definições importantes para se chegar ao problema da Descoberta de Processo, elas se encontram no livro *Process Mining: Data Science in Action* [26].

Redes de Petri

Em Mineração de Processos, muitos algoritmos utilizam Redes de Petri como uma linguagem de descrição do processo. Segue sua definição formal:

Definição 2.1 (*Redes de Petri*) Uma **Rede de Petri** é uma tripla $N = (P, T, F)$ onde P é um conjunto finito de **estados**, T é um conjunto finito de **transições** tais que $P \cap T = \emptyset$, e $F \subseteq (P \times T) \cup (T \times P)$ é um conjunto de arcos diretos, chamado de **fluxo da relação**. Uma Rede de Petri **marcada** é um par (N, M) , onde $N = (P, T, F)$ é uma Rede de Petri e onde $M \in \mathbb{B}(P)$ é um **multi conjunto** sobre P denotando a **marcação** da rede. O conjunto de todas as Redes de Petri marcadas é denotado $\mathcal{N}$.

Em geral, os modelos de processos não são Redes de Petri simples, mas, incluem também outros atributos. Considere também uma Rede de Petri rotulada.

Definição 2.2 (*Rede de Petri rotulada*) Uma **Rede de Petri rotulada** é uma tupla $N = (P, T, F, A, l)$ em que (P, T, F) é uma Rede de Petri, $A \subseteq \mathcal{A}$ é um conjunto de **atividades rotuladas**, e $l \in T \rightarrow A$ é uma **função de rotulação**.

Redes de Workflow (*WF-net*)

Dentre as várias extensões e simplificações da Rede de Petri, a rede de workflow (WF-net) é um subconjunto de Redes de Petri que é destacada pelas definições a seguir.

Definição 2.3 (*Transição de entrada/saída*) Seja $N = (P, T, F)$ uma Rede de Petri. Elementos de $P \cup T$ são chamados **nós**. Um nó x é um **nó de entrada** de outro nó y se e somente se existir um arco direto entre x e y , (p.e., $(x, y) \in F$). O nó x é um **nó de saída** de y se e somente se $(x, y) \in F$. Para qualquer $x \in P \cup T$, $\bullet x = \{y | (y, x) \in F\}$ e $x \bullet = \{y | (x, y) \in F\}$.

Definição 2.4 (*Regra de ativação*) Seja (N, M) uma Rede de Petri marcada com $N = (P, T, F)$ e $M \in \mathbb{B}(P)$. A transição $t \in T$ está **ativa**, denotada $(N, M)[t]$, se e apenas se $\bullet t \leq M$. A **regra de ativação** $[_\bullet]_\bullet \subseteq \mathcal{N} \times T \times \mathcal{N}$ é a menor relação que satisfaz qualquer $(N, M) \in \mathcal{N}$ e qualquer $t \in T$, $(N, M)[t] \implies (N, M)[t](N, (M \setminus \bullet t) \uplus t \bullet)$.

Definição 2.5 (Redes de workflow - WF-net) Seja $N = (P, T, F, A, l)$ uma Rede de Petri rotulada e $\bar{i}$ um identificador novo fora de $P \cup T$. N é uma **rede de workflow** (WF-net) se e somente se (a) P tem um estado de entrada i (estado inicial) tal que $\bullet i = \emptyset$, (b) P contém um estado de saída o (sumidouro) tal que $o \bullet = \emptyset$, e (c) $\bar{N} = (P, T \cup \{\bar{i}\}, F \cup \{(o, \bar{i}), (\bar{i}, i)\}, A \cup \{\tau\}, l \cup \{(\bar{i}, \tau)\})$ é fortemente conectado, p.e., existe um caminho direto entre qualquer par de nós em $\bar{N}$

Definições da Descoberta de Processos

Definição 2.6 (Problema geral de Descoberta de Processos) Seja L um registro de evento como definido acima. Um **algoritmo de Descoberta de Processos** é uma função que mapeia L para um modelo de processo tal que o modelo 'represente' o comportamento encontrado nos registros de eventos. O desafio é encontrar tal algoritmo.

Essa definição geral não apresenta qualquer característica ou propriedade esperada de um modelo gerado pela Descoberta de Processos, por isso, é importante desenvolver uma definição mais específica. Seguem conceitos introdutórios.

Definição 2.7 (Solidez) Seja $N = (P, T, F, A, l)$ uma WF-net com estado inicial i e sumidouro o . N é **sólido** se e somente se:

Segurança: $(N, [i])$ é seguro, p.e., estados não possuem mais de um marcador ao mesmo tempo.

Conclusão adequada: para cada marcador $M \in [N, [i]]$, $o \in M$ implica que $M = [o]$.

Opção de completar: para cada marcador $M \in [N, [i]]$, $[o] \in [N, M]$.

Ausência de partes mortas: $(N, [i])$ contém nenhuma transição morta (p.e., para qualquer $t \in T$, existe uma sequência habilitadora de ativação t).

Definição 2.8 (Caso, sequência, registro de evento) Seja $\mathcal{L}$ o **universo de casos**, p.e., o conjunto de todos os possíveis identificadores. Casos, como eventos, têm atributos. Para cada caso $c \in \mathcal{L}$ e nome $n \in AN$: $\#_n(c)$ é o valor do atributo n para o caso c ($\#_n(c) = \perp$ se o caso c não tiver o atributo n). Cada caso tem um atributo obrigatório chamado traço: $\#_{trace}(c) \in \mathcal{E}^*$. $\hat{c} = \#_{trace}(c)$ é um atalho para se referir ao traço de um caso.

Um **traço** é uma sequência finita de eventos $\sigma \in \mathcal{E}^*$ tal que cada evento aparece apenas uma vez, p.e., para $1 \leq i < j \leq |\sigma|$: $\sigma(i) \neq \sigma(j)$.

Um **registro de evento** é um conjunto de casos $L \subseteq \mathcal{L}$ tal que cada evento aparece pelo menos uma vez no registro inteiro, p.e., para cada $c_1, c_2 \in L$ tal que $c_1 \neq c_2$: $\partial_{set}(\hat{c}_1) \cap \partial_{set}(\hat{c}_2) = \emptyset$.

Definição 2.9 (Simples registro de evento) Seja $\mathcal{A}$ um conjunto de nomes de atividades. Um **simples traço** σ é uma sequência de atividade, p.e., $\sigma \in \mathcal{A}^*$. Um registro de evento **simples** L é um multi conjunto de traços em $\mathcal{A}$, p.e., $L \in \mathbb{B}(\mathcal{A}^*)$.

Por fim, segue a definição específica.

Definição 2.10 (*Problema específico de Descoberta de Processos*) Um **algoritmo de Descoberta de Processos** é uma função γ que mapeia um registro $L \in \mathbb{B}(\mathcal{A}^*)$ em uma Rede de Petri marcada $\gamma(L) = (N, M)$. Idealmente, N é uma WF-net **sólida** e todos os traços em L correspondem a possíveis sequências de ativação em (N, M) .

Neste trabalho, quando for citado apenas algoritmo, entenda-se por isso um algoritmo de Descoberta de Processos.

Características dos Algoritmos de Descoberta de Processo

A respeito dos algoritmos de Descoberta de Processos, uma maneira de entender sua capacidade é através de algumas características. A primeira e talvez mais importante é o **viés representacional**, que define a classe do modelo de processo que pode ser descoberto. O viés representacional determina o espaço de busca e potencialmente limita a expressividade do modelo descoberto. Dentre alguns exemplos de padrões que podem ou não ser reconhecidos estão: concorrência, laços, atividades silenciosas, ações duplicadas, divisão/junção do tipo OU, comportamento de livre escolha e hierarquia [26].

Outras características são: a capacidade de lidar com ruído no meio dos registros; a pressuposição de que os registros utilizados estão completos e apresentam todas as possibilidades de decisões de um processo; e a abordagem utilizada [26].

CrITÉrios de Qualidade de um Modelo de Processo

O modelo de processo resultante da Descoberta de Processos pode ser avaliado quanto a alguns critérios de qualidade, [26]. Existem quatro principais dimensões de qualidade que são mensuráveis e competem entre si:

1. **Aptidão (*Fitness*)** - capacidade do modelo de processo reproduzir novamente os registros de evento;
2. **Simplicidade (*Simplicity*)** - escolha do modelo mais simples que explique o comportamento dos registros de evento;
3. **Precisão (*Precision*)** - capacidade do modelo de não permitir comportamento não visto nos registros de evento;
4. **Generalização (*Generalization*)** - capacidade do modelo de permitir comportamento não observado.

2.1.2 Checagem de Conformidade (*Conformance Checking*)

A Checagem de Conformidade relaciona eventos no registro de eventos com atividades no modelo de processo e os compara. O objetivo é encontrar o que é comum e o que está em discrepância entre os comportamentos modelado e observado [26]. As métricas também se relacionam com os critérios de qualidade de um modelo de processo, sendo sempre associadas a um ou outro desses critérios.

A principal técnica utilizada, por todas as métricas, é a *repetição*, que é o uso de cada caso dos registros de eventos para verificar se existe uma transição correspondente na atividade marcada, sendo a atividade inicial marcada logo no começo, e a cada transição existente, a marcação segue para a próxima atividade. A mais simples métrica de Checagem de Conformidade usa os marcadores da repetição para fornecer a métrica conhecida como **precisão** [26].

Definição 2.11 (*Precisão*) Considere a técnica de repetição. Ao final de sua execução existirão quatro marcadores: **p** (marcadores produzidos), **c** (marcadores consumidos), **m** (marcadores ausentes) e **r** (marcadores que sobraram). O cálculo da precisão de um traço se dá por:

$$fitness(\sigma, N) = \frac{1}{2} \left(1 - \frac{m}{c} \right) + \frac{1}{2} \left(1 - \frac{r}{p} \right)$$

O cálculo completo sobre os registros de eventos se dá por:

$$fitness(L, N) = \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times m_{N, \sigma}}{\sum_{\sigma \in L} L(\sigma) \times c_{N, \sigma}} \right) + \frac{1}{2} \left(1 - \frac{\sum_{\sigma \in L} L(\sigma) \times r_{N, \sigma}}{\sum_{\sigma \in L} L(\sigma) \times p_{N, \sigma}} \right)$$

onde, $p_{N, \sigma}$, $c_{N, \sigma}$, $m_{N, \sigma}$ e $r_{N, \sigma}$, são os respectivos marcadores de um traço e $L(\sigma)$ é a frequência que um traço aparece nos registros de eventos.

2.2 Aprendizado de Máquina (*Machine Learning*)

Aprendizado de Máquina, de acordo com Alpaydin [1] é a programação de computadores para otimizar um critério de desempenho usando dados de exemplo ou experiência passada. Temos um modelo definido até alguns parâmetros, e aprendizado é a execução de um programa de computador para otimizar os parâmetros do modelo usando os dados de treinamento ou experiência passada. O modelo pode ser *preditivo* para fazer previsões no futuro, ou *descritivo* para ganhar conhecimento a partir de dados, ou ambos.

Aprendizado de Máquina usa a teoria da estatística na construção de modelos matemáticos, porque a tarefa central é fazer inferência a partir de uma amostra. O papel da ciência da computação é duplo: em primeiro lugar, no treinamento, precisamos de algoritmos eficientes para resolver o problema de otimização, bem como para armazenar

e processar a grande quantidade de dados que geralmente temos. Segundo, uma vez que um modelo é aprendido, sua representação e solução algorítmica para inferência precisa ser eficiente também. Em certas aplicações, a eficiência do algoritmo de aprendizado ou inferência, ou seja, sua complexidade de espaço e tempo, pode ser tão importante quanto sua precisão preditiva. Considere as seguintes aplicações de Aprendizado de Máquina que serão abordados ao longo deste trabalho.

2.2.1 Agrupamentos (*Clustering*)

Agrupamentos é uma técnica de Aprendizado de Máquina não-supervisionada de encontrar padrões nos dados, isto é, estes algoritmos funcionam sem atributos de classe [16].

O algoritmo **Expectation-Maximization (EM)** é usado na estimativa de máxima semelhança onde o problema envolve dois conjuntos de variáveis aleatórias, das quais uma, X , é observável e a outra, Z , está oculta. O objetivo do algoritmo é encontrar o vetor de parâmetro ϕ que maximiza a probabilidade de os valores observados de X , $L(\phi|X)$. Mas nos casos em que isso não é viável, associamos as variáveis ocultas extra Z e expressamos o modelo subjacente usando ambos, para maximizar a probabilidade da distribuição conjunta de X e Z , a probabilidade completa $L_c(\phi|X, Z)$.

2.2.2 Seleção de Atributos (*Selecting attributes*)

A Seleção de Atributos envolve a busca através de todas as combinações possíveis de atributos nos dados para encontrar qual subconjunto de atributos funciona melhor para a previsão [16]. Para fazer isso, dois objetos devem ser configurados: um avaliador de atributo e um método de pesquisa. O avaliador determina qual método é usado para atribuir um valor a cada subconjunto de atributos. O método de pesquisa determina qual estilo de pesquisa é executado.

Revisão bibliográfica

A revisão bibliográfica é resultado de uma pesquisa a partir da citação dos principais algoritmos de Descoberta de Processos identificados em van der Aalst [24]. O objetivo foi de verificar a evolução dos principais algoritmos, realizar comparações entre eles e encontrar ambientes de classificação.

3.1 Algoritmos de Descoberta de Processo

Como visto na Seção 2.1.1, os algoritmos geram modelos com certas características que permitem compara-los. A maneira mais comum de classificação consiste no uso da abordagem computacional [24].

3.1.1 Famílias de Abordagem Computacional

A divisão mais conhecida dos algoritmos se dá em famílias pela abordagem computacional utilizada por ele. Na pesquisa exploratória realizada, foram encontrados vários algoritmos que se enquadram nas seguintes famílias:

1. **Algorítmica direta** que dentre essa família estão os algoritmos α [29], $\alpha+$ [10], α^* [18], $\alpha\#$ [40], $\alpha++$ [39], β [41] e os algoritmos de mineração heurística HeuristicsMiner [38], HeuristicsMiner++ [8], FuzzyMiner [15], FlexibleHeuristicsMiner [37] e Fodina [33];
2. **Em duas fases** que dentre essa família estão algoritmos de mineração baseada em regiões [25], [3], ILP Miner [30], InductiveMiner [17], HybridILPMiner [43];
3. **De inteligência computacional** que dentre essa família estão os algoritmos Genético [11], ETM [5] e o ProDiGen [35];
4. **Parcial** que dentre essa família estão algoritmos que fazem uso de linguagem declarativa [27], Declare Miner [19], MINERful++ [14], UnconstrainedMiner [42];

5. E existem também algoritmos que utilizam combinações dessas abordagens, como o FusionMiner [12].

Dessa classificação, observa-se a presença de dois grandes grupos na abordagem algorítmica direta, separados pelo uso de computação heurística. Algo semelhante ocorre na abordagem de duas fases, sendo a maneira que o algoritmo realiza aproximações para separar as atividades e gerar os caminhos diferente, de acordo com cada um, e podem não ter relação, mesmo que a abordagem seja parecida. Além disso, a abordagem parcial considera um tipo de resultado diferente dos demais, sendo declarativo o modelo gerado. Isso por si só impede a comparação dos modelos gerados com os demais algoritmos das outras famílias, pois não há conversão compatível.

Considerando-se os algoritmos mais recentes, segue Tabela 3.1 comparativa de como eles são apresentados quanto aos dados utilizados nos testes, nas comparações com outros algoritmos e nas métricas utilizadas. As métricas serão apresentadas posteriormente neste trabalho 3.2. Observe que os dados usados fazem referência a pelo menos um trabalho anterior, e que o caso mais comum é usarem dados gerados, ou seja, que não representam um caso real existente. Outra observação que se destaca é a falta de padronização na escolha das métricas aplicadas nas comparações, havendo pouquíssima coincidência delas entre os trabalhos. Além disso, percebe-se o uso de algoritmos a mais tempo na literatura para comparação, como o Heuristics Miner e o Declare Miner.

Algoritmo	Dados usados	Algoritmos comparados	Métricas aplicadas
Fodina, 2014 [33]	Dados gerados [21], randômico, permutado e casos reais	a++, Heuristics Miner, Flexible Heuristics Miner	<i>Average Alignment Based Trace Fitness, One Align Precision, Best Align Precision, ETC Precision, Behavioral Recall, Weighted Behavioral Precision, Weighted Behavioral Generalization, Number of Arcs, Places and Transitions</i>
Inductive Miner, 2013 [17]	Dados gerados [24]	-	-
Hybrid ILP Miner, 2015 [43]	Dados gerados, em especial [20]	-	<i>Replay-fitness, precision</i>
ProDiGen, 2014 [35]	Dados gerados [9] e reais [6]	Heuristics Miner, Genetic Miner	<i>Behavioral precision, Behavioral recall, Structural precision, Structural recall, Proper completion, alignment precision, weighted P/T average arc degree</i>
MINERful++, 2013 [14]	Dados gerados e reais [32, 31]	Declare Miner	-
Unconstrained Miner, 2013 [42]	Dados reais [32]	Declare Miner, MINERful++	-
FusionMiner, 2015 [12]	Dados gerados e reais [13]	Heuristics Miner, ILP Miner, DeclareMiner	<i>Behavioral Precision</i>

Tabela 3.1: Comparativo da apresentação de algoritmos selecionados.

3.2 Métricas de Checagem de Conformidade

A etapa de Checagem de Conformidade trabalha com métricas que captam os critérios de qualidade de um modelo de processo. As métricas foram revisadas dos trabalhos de Blum et al [4], Van den Broucke et al [34] e dos algoritmos Fodina [33], Hybrid ILP Miner [43], ProDiGen [35] e FusionMiner [12] e cobrem os quatro critérios de qualidade de um modelo de processo. Segue a compilação na Tabela 3.2.

Métrica	Fontes	Aptidão	Simplicidade	Precisão	Generalização
Advanced Behavioural Appropriateness	[4], [34]			X	
Advanced Structural Appropriateness	[4], [34]		X		
Alignment Based Fitness	[34]	X			
Alignment Based Precision	[34]			X	
Alignment Based Probabilistic Generalization	[34]				X
Alignment Precision	[35]			X	
Average Alignment Based Trace Fitness	[34], [33]	X			
Average Node Arc Degree	[34]		X		
Behavioral Precision	[4], [35], [12]			X	
Behavioral Profile Conformance	[34]	X			
Behavioral Recall	[4], [34], [33], [35]	X			X
Behavioral Similarity	[4]			X	X
Best Align Precision	[34], [33]			X	
Count Arcs	[34], [33]	X			
Count Cut Vertices	[34]	X			
Count Nodes	[34]	X			
Count Places	[34], [33]	X			
Count Transitions	[34], [33]	X			
ETC Conformance	[34]			X	
ETC Precision	[4], [34], [33]			X	
Fitness	[4], [34]	X			
Negative Event Generalization	[34]				X
Negative Event Precision	[34]			X	
Negative Event Recall	[34]	X			
One Align Precision	[34], [33]	X			
Parsing Measure	[4]	X			
Partial Fitness Complete	[4]	X			
Precision	[4], [43]			X	X
Proper Completion	[34], [35]	X			
Replay-fitness	[43]	X			
Simple Behavioural Appropriateness	[34]			X	
Simple Structural Appropriateness	[34]		X		
Simplicity	[4]		X		
Specificity	[4]			X	
Structural precision	[4], [35]			X	
Structural recall	[4], [35]				X
Structural Similarity	[4]			X	X
Successful Execution	[34]	X			
Weighed P/T Average Arc Degree	[35]	X			
Weighted Behavioral Generalization	[33]				X
Weighted Behavioral Precision	[33]			X	

Tabela 3.2: Compilação de métricas e sua relação com os critérios de qualidade [4], [34], [33], [43], [35] e [12].

3.3 Modelo de Ambiente de Avaliação e Comparação de Algoritmos

Com o avanço da Mineração de Processos, ferramentas de classificação fazem-se necessárias para medir as diferenças e melhorias na área. Em especial, para a Descoberta de Processos, temos um *Modelo* por Rozinat et al [23]. Esse modelo de ambiente de classificação considera o uso iterativo de uma maneira de avaliação e comparação com a Descoberta de Processos e simulação dos modelos gerados, podendo ser visto na Figura 3.1.

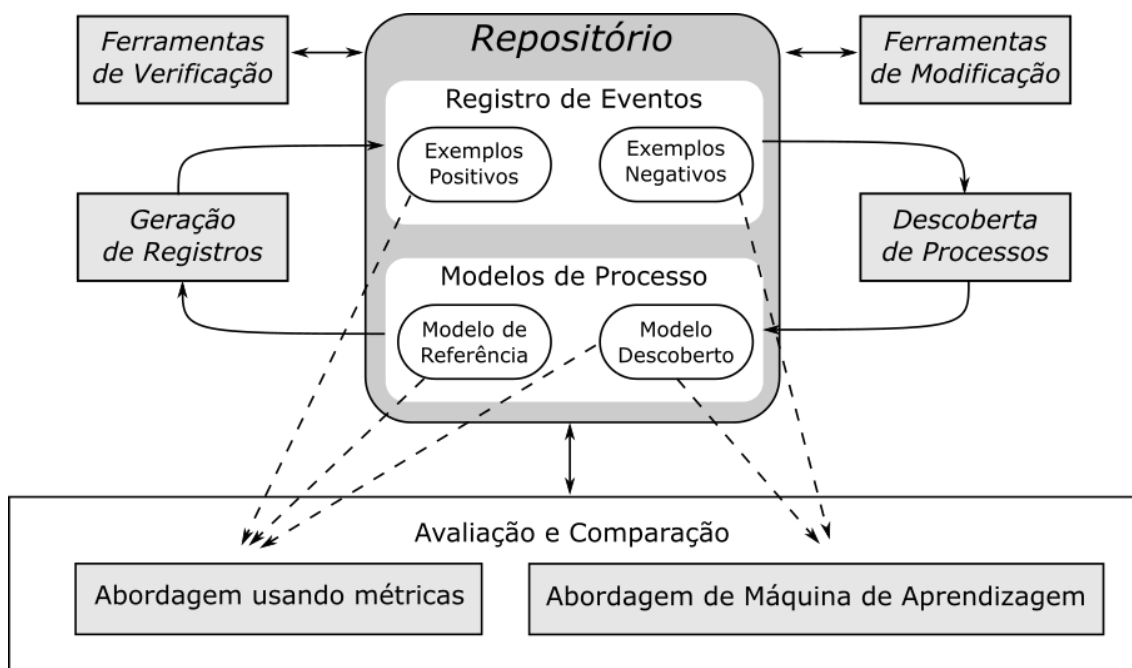


Figura 3.1: Ambiente modelo, apresentado por Rozinat et al [23].

No modelo proposto, observa-se, do contexto de Mineração de Processos, as fases de Descoberta e Checagem de Conformidade bem presentes, compondo um repositório onde serão mantidos dados para o histórico de recomendações. Elas são vistas na caixa de *Descoberta de Processos* à direita e na *Avaliação e Comparação* na base da figura.

O *Repositório*, parte central, mantém os registros de eventos, tanto em exemplos positivos quanto negativos e os modelos de processo, tanto os de referência quanto os descobertos. Ou seja, é responsável por manter o histórico das avaliações já feitas. As interações com o Repositório se dão por meio de ferramentas de verificação e modificação, que manipulam os registros e os modelos, como vistos nas caixas ao topo *Ferramentas de Verificação* e *Ferramentas de Modificação*.

Existe uma iteratividade de Descoberta de Processos, com os registros de eventos existentes, que serve para gerar novos modelos, e a partir destes, gerar novos registros.

Esse ciclo serve para gerar os dados necessários que serão usados na avaliação e comparação, de forma que um modelo de referência inicialmente é usado com possíveis registros existentes ou gerados pelo arcabouço. Os algoritmos de Descoberta de Processos são executados sobre esses registros, gerando novos modelos, que, então, são comparados com o de referência de alguma forma.

A existência de dois tipos de registros de eventos se dá para o uso nas duas abordagens possíveis de avaliação e comparação. A abordagem usando métricas requer exemplos positivos, ou seja, de dados que foram de fato gerados pelo modelo. Já a abordagem de Aprendizado de Máquina requer exemplos negativos, ou seja, realiza uma comparação com os modelos descobertos para invalidá-los.

3.4 Ambiente Implementado por Wang et al

Uma das implementações do ambiente modelo 3.3 foi apresentado por Wang et al [36]. Ela propõe um meio de avaliar a qualidade de um algoritmo com base em semelhanças estruturais e comportamentais usando a abordagem de Aprendizado de Máquina.

Na Figura 3.2, a proposta de Wang et al é apresentada e pode-se observar os componentes do ambiente modelo na sua descrição. O trabalho desenvolvido foi de utilizar uma extração de recursos dos modelos de referência e reduzir esses recursos para fazer uma análise de regressão, com o objetivo de prever o melhor algoritmo nos modelos corporativos de entrada do arcabouço.

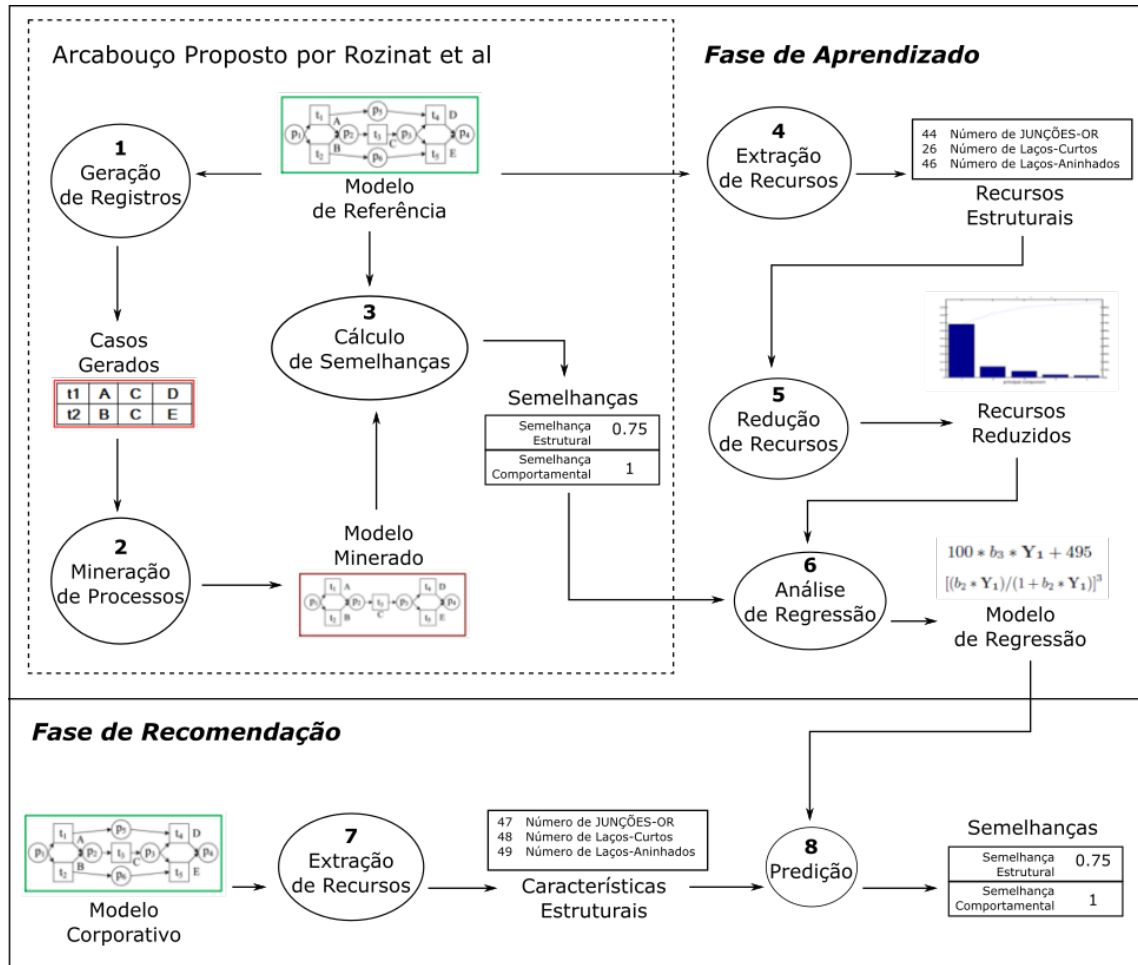


Figura 3.2: Ambiente implementado por Wang et al [36], com acréscimo de comparação de recursos e modelos de redução.

Conforme apontado pelos próprios autores, existe um problema de tempo para realizar a avaliação e retornar um resultado. Por isso, eles propõem uma maneira de regredir o modelo avaliado para realizar aproximações e tornar mais rápida a predição. Mas, os resultados não foram satisfatórios, principalmente para a análise estrutural, ficando aquém em mais de 50% do valor empírico.

3.5 Ambiente Implementado por Ribeiro et al

Outra implementação do ambiente modelo 3.3 é realizada por Ribeiro et al [22], que apresenta um sistema com abordagem de Aprendizado de Máquina, assim como Wang et al. Na Figura 3.3, está a representação desse trabalho, tendo o ciclo inicial apresentado por Rozinat et al, de um repositório com modelos de processo e registros de evento iterando com a Descoberta de Processo e verificação de conformidade. A inovação apresentada encontra-se nesse repositório com o acréscimo de *Resultados Experimentais*,

que são métricas levantadas da Descoberta de Processo e da Checagem de Conformidade usando modelos genéricos acumulados ao longo do tempo da execução do ambiente.

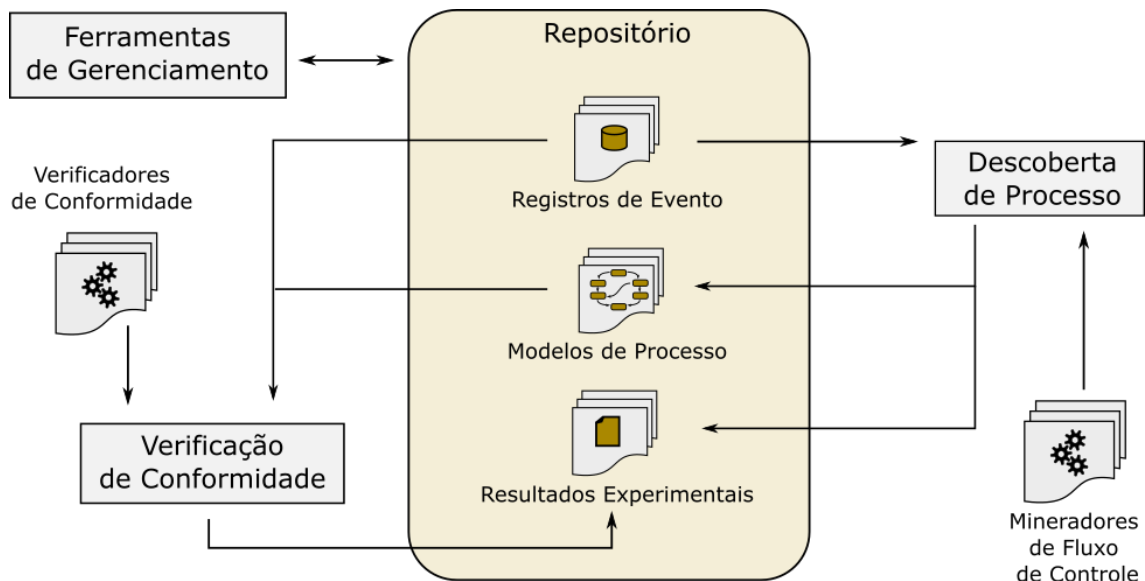


Figura 3.3: Ambiente implementado por Ribeiro et al [22], com histórico de comparações e avaliações já realizadas para sugestão, parte 1.

Como se vê na Figura 3.4, diferentemente de Wang et al, Ribeiro et al faz análises além da estrutural e comportamental, e, ao invés de realizar redução dos recursos, faz-se a geração de modelos de previsão. Igualmente ao que já foi proposto, existe uma extração de recursos do repositório e da entrada fornecida ao ambiente. Havendo, assim, uma classificação do repositório com base nesses recursos e a posterior previsão utilizando os recursos extraídos da entrada.

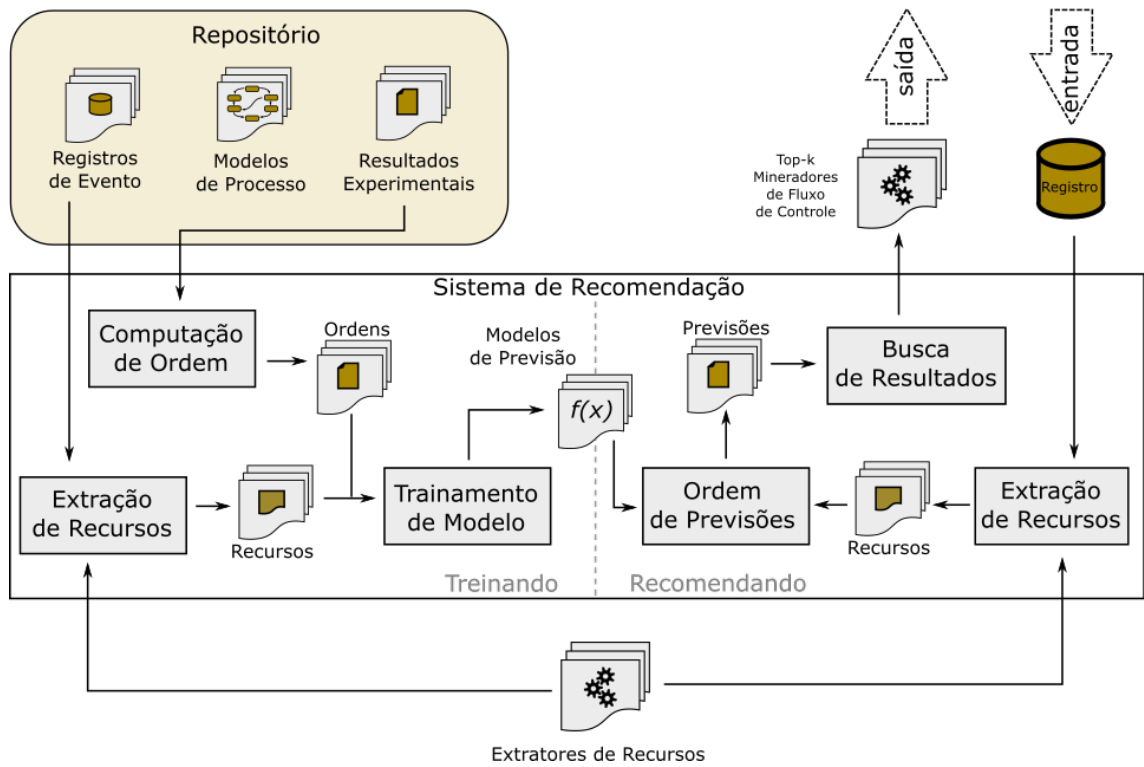


Figura 3.4: Ambiente implementado por Ribeiro et al [22], com histórico de comparações e avaliações já realizadas para sugestão, parte 2.

Materiais e Métodos

Neste capítulo estão a metodologia utilizada para desenvolver a pesquisa e as ferramentas utilizadas.

4.1 Metodologia

Esta pesquisa exploratória iniciou-se pelo estudo dos algoritmos apresentados no livro de Mineração de Processo [24] e suas referências mais citadas. Com isso, foi possível encontrar outros algoritmos e observar a evolução das famílias algorítmicas.

A observação dessa evolução trouxe à tona a falta de padrões de comparação dos algoritmos, como revelado na Tabela 3.1. Essa constatação levou à procura de ambientes de comparação de algoritmos, descritos nas Seções 3.3, 3.4 e 3.5.

Como visto no ambiente modelo na Seção 3.3, uma forma de classificar os algoritmos de Descoberta de Processos é gerando modelos de processo a partir de registros de eventos, e então compará-los usando como critérios de qualidade as métricas de Checagem de Conformidade. Entretanto, esse modelo apenas define o fluxo para se executar todas as etapas necessárias para realizar a classificação, sem explicitar ao certo quais os parâmetros em cada etapa, seja processos de negócio, registros de eventos ou métricas.

Este trabalho atua na explicitação dos parâmetros em 4 etapas. O ambiente proposto pode ser ilustrado como segue, na Figura 4.1.

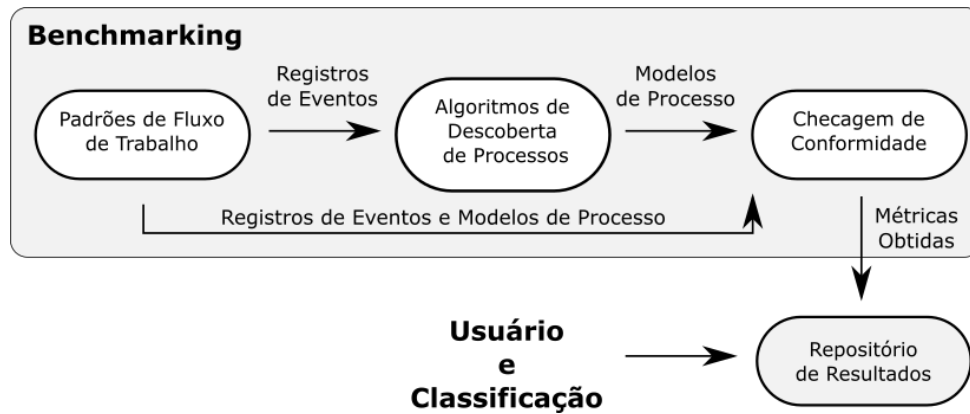


Figura 4.1: Ambiente de classificação de algoritmos utilizado.

1. Estabelecimento de **Padrões de Fluxo de Trabalho** a serem usados para gerar os registros de eventos;
2. Utilização dos registros de eventos nos **Algoritmos de Descoberta de Processos** a serem avaliados;
3. Aplicação de métricas de **Checagem de Conformidade** sobre os modelos descobertos;
4. Realização de agrupamentos e classificação sobre as métricas levantadas para gerar o **Repositório de Resultados**.

Como foi constatado na Tabela 3.1, não há ainda para a área de Mineração de Processos uma maneira padrão de comparação entre os algoritmos de Descoberta de Processos. No ambiente proposto isso fica destacado como sendo um *Benchmarking* disponibilizado que supre essa carência. Ao executar os passos da Mineração de Processos, utilizando como início os padrões de fluxo de trabalho definidos, o *Repositório de Resultados* é acrescido dos resultados. Esses resultados auxiliam a escolha do usuário para o melhor algoritmo para dada situação e permite a classificação dos algoritmos em grupos por funcionalidade apresentada.

(1) Inicialmente, da observação do modelo estendido por Ribeiro et al 3.5, apresentado na Seção 3.5, aproveita-se o uso de registros de eventos iniciais e a abstração dos recursos. Com base nisso, é estabelecido para o ambiente proposto o uso de registros de eventos gerados de modelos que seguem *padrões de fluxo de trabalho*, de modo a observar o comportamento de cada algoritmo para cada um deles. Diferentemente do ambiente anterior, o ambiente proposto neste trabalho não armazenaria os modelos de entrada no arcabouço, mas teria modelos com esses padrões que podem ser atualizados e aumentados com o tempo.

Esses modelos de processo com os padrões de fluxo de trabalho são gerados na ferramenta PLG2 [7]. Nele já estão definidos os padrões de fluxo de trabalho, bastando

gerar o modelo de processo e os registros de eventos a serem utilizados na Descoberta de Processos e comparações na Checagem de Conformidade.

(2) Em seguida, os registros de eventos gerados são utilizados em cada *algoritmo de Descoberta de Processo* selecionado, como descrito na Seção 4.2.1.

(3) Os modelos gerados por eles são, então, mensurados com as métricas descritas na Seção 3.2, em especial as disponíveis no CoBeFra [34] de *Checagem de Conformidade*. Apenas dessas métricas deve-se poder escolher o melhor algoritmo para cada padrão de fluxo de trabalho.

(4) Ao final, para realizar a análise dos dados e conseguir classificar os algoritmos, realizam-se agrupamentos e outras análises com as métricas levantadas, como apresentado na Seção 4.2.4, para encontrar grupos de algoritmos e validar uma nova maneira de agrupá-los, compondo assim o *repositório de resultados*.

4.2 Ferramentas

Nesta seção são apresentadas as ferramentas utilizadas.

4.2.1 ProM 6.6

A contribuição pública mais citada da área de Mineração de Processos é a ferramenta **Process Mining Framework - ProM 6.6** [26], que consiste em uma aplicação java de código livre extensível que dá suporte a vários complementos.

O objetivo do grupo que a mantém é fazer dela a ferramenta padrão para a comunidade acadêmica, ao manter ativos os contribuintes e usuários reconhecidos, criando conhecimento para desenvolver a tecnologia de Mineração de Processos.

A visão deles é avançar ativamente o estado da arte na tecnologia de Mineração de Processos ao desenvolver métodos que realmente funcionam, ao criar comunidades abertas e ao prover uma plataforma estável e facilmente extensível, que otimizadamente apoia a disciplina.

A plataforma trabalha com interfaces de integração de modo que, ao importar um objeto para dentro dela, ele se torna disponível para uso dos complementos instalados. E cada resultado obtido, gera objetos que são automaticamente integrados à plataforma e, em alguns casos, um relatório da execução através de um gráfico, estatística ou outros.

Complementos no ProM 6.6 de Descoberta de Processos

Nos trabalhos dos algoritmos de Descoberta de Processos, fica claro que o ProM é a plataforma padrão para a implementação dos mesmos. Mas, seja por evolução ou falta

de manutenção nos pacotes individuais, boa parte dos algoritmos não se encontra mais disponível na versão 6.6 da plataforma.

Em função disso, foram selecionados alguns dos algoritmos supracitados para utilização neste trabalho conforme mostra a Tabela 4.1. Vale ressaltar que os complementos de métricas trabalham exclusivamente com Redes de Petri e, por isso, a conversão de outro formato para este é requerido. Nos casos em que foi necessário realizar este procedimento, o nome do complemento conversor também se encontra apresentado.

Complemento	Autor	Algoritmos Utilizados
<i>Alpha Miner</i>	Van Zelst, van Dongen	Alpha, Alpha#
<i>Mine for a Heuristics Net using Heuristics Miner</i>	Weijters	Heuristics Miner
<i>Convert Heuristics net into Petri net</i>	Ribeiro	Conversor de rede heurística para Rede de Petri
<i>Mine for a Petri Net using ILP</i>	Van der Wiel	ILP
<i>ILP-Based Process Discovery</i>	Van Zelst	Hybrid-ILP
<i>Mine a Process Tree with ETMd</i>	Buijjs	Evolutionary Tree Miner (ETM)
<i>Convert Process Tree to Petri Net</i>	Leemans	Conversor de árvore de processo para Rede de Petri

Tabela 4.1: Algoritmos e complementos utilizados do ProM 6.6 para Descoberta de Processos.

4.2.2 PLG2

A aplicação **Processes and Logs Generator 2 - PLG2** é trabalho do Burattin [7], e também consiste em uma aplicação java de código livre.

Trata-se em um gerador de modelos de processos, que na versão mais recente (v2.0.5 - 18 de Novembro de 2016), exporta o modelo gerado nos mais diferentes formatos, em especial a Rede de Petri. Para essa geração, ele disponibiliza várias características que envolvem a escolha da profundidade, quantidade máxima de desvios (E ou OU exclusivo), peso para gerar sequências maiores, peso para gerar atividades sem sequência, peso para atividades que são puladas, peso de desvios E e OU exclusivo, peso para laços e a probabilidade para gerar objetos de dados. Com essas configurações, o modelo é aleatoriamente construído.

Essas propriedades são padrões de fluxo de controle já conhecidos e mapeados por van der Aalst em 2003 [28], sendo eles (a) WCP-1: sucessão direta de duas atividades; (b) WCP-2: execução paralela; (c) WCP-3: sincronização de ramos paralelos; (d) WCP-4: execução mútua; (e) WCP-5: convergência de ramos; (f) WCP-21: habilidade de executar subprocessos repetidamente.

Além disso, essa ferramenta permite exportar registros de eventos como uma simulação de possíveis execuções do processo. Para tal, ele conta com outras opções, como ruído (em várias situações, como nos nomes, nas sequências, nos objetos de dados),

quantidade de traços, probabilidades diversas no caso de ruído e tamanho máximo de início e fim do traço. Essa exportação é feita em formatos aceitos no ProM 6.6.

4.2.3 CoBeFra

Dentro da área de pesquisa de Mineração de Processos, duas tarefas proeminentes podem ser discernidas. Primeiro, a descoberta de processos lida com a construção automática de um modelo de processo a partir de um registro de eventos. Em segundo lugar, a verificação da conformidade se concentra na avaliação da qualidade de um modelo de processo descoberto ou projetado em relação ao comportamento real capturado nos registros de eventos.

Em relação à verificação de conformidade, várias técnicas e métricas foram desenvolvidas e descritas na literatura. No entanto, o domínio de Mineração de Processos carece de um quadro abrangente para avaliar a validade de um modelo de processo a partir de uma perspectiva quantitativa.

Com *CoBeFra* [34], oferece-se a arquitetura para uma estrutura extensível dentro do ProM, permitindo o cálculo consistente, comparativo e repetível de métricas de conformidade. Para o desenvolvimento e avaliação tanto de descoberta de processo como de técnicas de conformidade, tal estrutura é muito valiosa.

Ele é desenvolvido em Java e suas métricas aplicam-se a Redes de Petri e registros de eventos. Como entrada, aceitam-se os resultados do Prom 6.6 e do PLG2.

Dele, as seguintes métricas foram utilizadas: *Advanced Behavioural Appropriateness*, *Advanced Structural Appropriateness*, *Alignment Based Fitness*, *Alignment Based Precision*, *Alignment Based Probabilistic Generalization*, *Average Node Arc Degree*, *Behavioral Profile Conformance*, *Best Align Precision*, *Count Arcs*, *Count Cut Vertices*, *Count Nodes*, *Count Places*, *Count Transitions*, *ETC Conformance*, *Fitness*, *Negative Event Generalization*, *Negative Event Precision*, *Negative Event Recall*, *One Align Precision*, *Proper Completion*, *Simple Behavioural Appropriateness*, *Simple Structural Appropriateness*, *Successful Execution*, *Weighed P/T Average Arc Degree*.

4.2.4 WEKA

A plataforma WEKA, na versão 3.8.1 [16], é uma coleção de algoritmos de Aprendizado de Máquina e ferramentas de pré-processamento. Ele provê suporte extensivo para todo processo em mineração de dados experimental, incluindo o preparo dos dados, avaliação estatística dos esquemas de aprendizado e visualização dos dados de entrada e os resultados do aprendizado. O sistema é de código aberto e escrito em java.

Para separar os algoritmos de Descoberta de Processos em grupos, foi utilizado o algoritmo de Agrupamento *Expectation-Maximization (EM)*. Ele dá uma distribuição

de probabilidade para cada instância que indica a probabilidade dela pertencer a um dos agrupamentos. **EM** pode decidir quantos agrupamentos criar por validação cruzada.

Além dele, foram utilizados componentes de Seleção de Atributos para validar a relevância das métricas geradas. Para esse fim, foi usado o componente **CfsSubsetEval** com **BestFirst**. Eles fazem uma validação em todo o conjunto de dados para identificar quão relevante é um atributo para classificar os dados.

Experimentos

Neste capítulo estão as configurações utilizadas nas ferramentas do arcabouço e os resultados experimentais obtidos.

5.1 Configurações

5.1.1 PLG2

O PLG2, que foi utilizado para gerar os modelos e os registros de eventos, foi manipulado para ter os modelos mais simples possíveis observando cada padrão por ele trabalhado, de modo que cada padrão tivesse um modelo que o representasse e outros derivados que aumentavam o tamanho do modelo, mas, utilizando ainda o mesmo padrão.

A respeito da geração dos registros de eventos, foram mantidos sem ruídos e com 1000 traços. Por se tratarem de modelos simples sem muitas opções de desvio no processo, não há necessidade de tantos traços, pois a quantidade de traços repetidos já é suficiente para se ter amostras de todo comportamento possível do modelo.

5.1.2 Algoritmos de Descoberta de Processos

Em geral, os algoritmos de Descoberta de Processos requerem um classificador dos eventos. Para este campo, sempre foi deixada a configuração padrão [*Event Name AND Lifecycle transition*].

Para o pacote do **Alpha Miner**, foi configurada a versão a ser utilizada, que para observar a diferença evolutiva, utilizou-se neste parâmetro a versão Alpha e em outra execução o Alpha#.

Para o pacote do **Heuristics Miner**, nenhuma configuração foi alterada, pois elas se referem, principalmente, a casos de dependência de longas distância. O padrão estava como segue: *Relative-to-best*: 5.0, *Dependency*: 90.0, *Length-one-loops*: 90.0, *Length-two-loops*: 90.0, *Long distance*: 90.0, *All tasks connected*: marcado, *Long distance dependency*: desmarcado, *Ignore loop dependency thresholds*: marcado.

Para o pacote do **ILP Miner**, existe configurações dos algoritmos que o implementam, com algumas variações e extensões disponíveis. A única alteração foi do *Number of places*, que para ela foi selecionada a opção *Before & After Transition*, pois para a checagem de métricas, essa opção faz a geração de um estado final, que influencia os resultados e sem ele pode até não ser bem processada a análise das métricas. O padrão estava como segue: *ILP Solver: Java-ILP & LpSolve 5.5*, *ILP Variant: Petri Net*, *Search for separete initial places*: marcado, *ILP Extensions*: todas desmarcadas.

Para o pacote do **ILP-Based Miner**, foi mantida a opção *Detailed configuration* desmarcada e não houve nada a ser configurado.

Para o pacote do **ETMd Miner**, foram mantidas as configurações padrão a exceção do número máximo de gerações. No entanto, ele permite configurar os pesos considerando várias opções relacionadas às métricas de qualidade, para influenciar na eleição dos modelos gerados a cada geração. O objetivo final deste algoritmo é procurar precisão próxima a 1.

Em especial ao ETMd, os experimentos variaram a geração máxima, esperando uma convergência dos resultados para então considerar o melhor deles. No caso de não haver convergência, aumentava-se o número de gerações para observar se haveria convergência nessas novas amostras. O número inicial de gerações era de 50, depois 100, depois 200 e por último 300. Em alguns casos, observou-se fitness máximo logo com 50 gerações e, para esses casos, houve redução na quantidade de gerações máxima para verificar se ainda haveria convergência, para eles, reduziu-se para 10 e até 1 geração. Como o ETMd tem como estratégia o uso de outros algoritmos de Pescoberta de Processos na sua primeira geração, pode ocorrer um máximo que persiste por todas as gerações.

Ao se utilizar o melhor valor encontrado para o máximo de gerações no ETMd, evita-se usar qualquer valor estatístico entre eles, pois para a mesma geração pode haver convergência em uma só métrica, mas, divergência nas demais. Como é um algoritmo genético, esperou-se observar pelo menos 2 resultados iguais para se ter certeza do melhor valor.

5.1.3 Algoritmos de Checagem de Conformidade

As métricas providas pelo CoBeFra já estão configuradas para uma execução padrão, e caso necessário, algumas delas permitem ajustar pesos ou alterar algoritmos como o da repetição. Como não foi utilizado ruído, não há necessidade de ficar configurando pesos e custos para a penalização de discrepâncias, ficando todas as métricas nas suas configurações padrão. Além disso, existe um mapeamento do modelo de entrada, Rede de Petri, para alinhar com o CoBeFra os nomes das atividades. Nesses mapeamentos, quando necessário, as atividades não mapeadas foram configuradas como sendo invisíveis. Essa

configuração foi necessária para o Heuristics Miner em todos os casos, para o ETMd em alguns casos e para o modelo original em alguns casos.

Ressalta-se que no decorrer da execução, algumas métricas retornavam 'Warning' ou 'Error', e isso por si só não inviabiliza o uso da mesma, pois é uma informação importante da aplicação da métrica. O fato de gerar um valor que possa ser comparado é que foi verificado, mesmo que ele seja zero.

5.1.4 WEKA

O componente clusterizador **EM**, foi mantido em sua configuração padrão, para automaticamente identificar a quantidade de clusters possíveis. Houve variação do atributo classificador, que para ele, variou-se para observar os resultados.

O mesmo aconteceu com os componentes de seleção de atributos, **CfsSubsetEval** com **BestFirst**, que ficaram na configuração padrão com o uso de todo o conjunto de dados para treino e o atributo 'Algoritmos' como critério avaliador.

5.2 Execução

Todos os modelos possuem exatamente uma entrada e saída. Para o algoritmo ETMd, o número apresentado reflete a melhor execução, como foi explicitado da sua estratégia. Para os números exibidos, foi feita uma aproximação de modo a todos estarem com o mesmo número de casas decimais.

5.2.1 Padrão de Sequência de Atividades

O primeiro padrão, em que foram executados os algoritmos de Descoberta de Processo e as métricas, foi na sequência de atividades. Como primeiro modelo, apenas duas atividades, como está na Figura 5.1. Os resultados obtidos estão na Tabela 5.1.



Figura 5.1: *Desenho do modelo de Processo 1, no padrão de sequência.*

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,66667	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	1,60000	1,60000	1,71429	1,60000	1,50000	1,60000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	4,00000	4,00000	4,00000	6,00000	4,00000	3,00000	4,00000
Count Cut Vertices	1,00000	1,00000	1,00000	2,00000	1,00000	1,00000	1,00000
Count Nodes	5,00000	5,00000	5,00000	7,00000	5,00000	4,00000	5,00000
Count Places	3,00000	3,00000	3,00000	4,00000	3,00000	2,00000	3,00000
Count Transitions	2,00000	2,00000	2,00000	3,00000	2,00000	2,00000	2,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,66667	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	1,00000	1,00000	1,00000	1,50000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,80000	0,80000	0,80000	0,57143	0,80000	1,00000	0,80000
Succesful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,66667	1,66667	1,66667	1,75000	1,66667	1,50000	1,66667

Tabela 5.1: Resultados do experimento no padrão de sequência conforme Figura 5.1.

Aplicou-se agora uma profundidade maior do padrão de atividades em sequência, tendo 5 atividades ao todo, como está na Figura 5.2. Os resultados obtidos estão na Tabela 5.2.

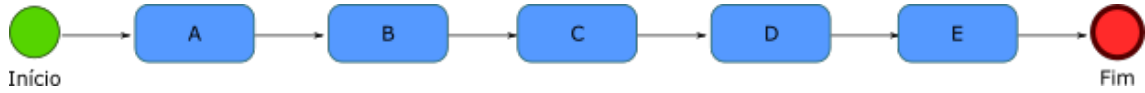


Figura 5.2: Desenho do modelo de Processo 1.1, no padrão de sequência.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,55556	0,75000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,81818	1,81818	1,81818	1,89474	1,88235	4,11111	1,81818
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	10,00000	10,00000	10,00000	18,00000	16,00000	37,00000	10,00000
Count Cut Vertices	4,00000	4,00000	4,00000	8,00000	7,00000	0,00000	4,00000
Count Nodes	11,00000	11,00000	11,00000	19,00000	17,00000	18,00000	11,00000
Count Places	6,00000	6,00000	6,00000	10,00000	9,00000	13,00000	6,00000
Count Transitions	5,00000	5,00000	5,00000	9,00000	8,00000	5,00000	5,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,33333	0,83333	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	1,00000	1,00000	1,00000	1,20000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,63636	0,63636	0,63636	0,36842	0,41176	0,38889	0,63636
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,83333	1,83333	1,83333	1,90000	1,88889	5,12308	1,83333

Tabela 5.2: Resultados do experimento no padrão de sequência conforme Figura 5.2.

Novamente, aplicou-se uma profundidade ainda maior do padrão de atividades em sequência, tendo 8 atividades ao todo, como está na Figura 5.3. Os resultados obtidos estão na Tabela 5.3.



Figura 5.3: Desenho do modelo de Processo 1.2, no padrão de sequência.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,53333	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,88235	1,88235	1,88235	1,93548	1,87500	5,28205	1,88235
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	16,00000	16,00000	16,00000	30,00000	15,00000	103,00000	16,00000
Count Cut Vertices	7,00000	7,00000	7,00000	14,00000	7,00000	0,00000	7,00000
Count Nodes	17,00000	17,00000	17,00000	31,00000	16,00000	39,00000	17,00000
Count Places	9,00000	9,00000	9,00000	16,00000	8,00000	31,00000	9,00000
Count Transitions	8,00000	8,00000	8,00000	15,00000	8,00000	8,00000	8,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,22222	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	1,00000	1,00000	1,00000	1,12500	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,58824	0,58824	0,58824	0,32258	0,62500	0,25641	0,58824
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,88889	1,88889	1,88889	1,93750	1,87500	8,09879	1,88889

Tabela 5.3: Resultados do experimento no padrão de sequência conforme Figura 5.3.

Uma última vez, aplicou-se maior profundidade do padrão de atividades em sequência, tendo 13 atividades ao todo, como está na Figura 5.4. Os resultados obtidos estão na Tabela 5.4.

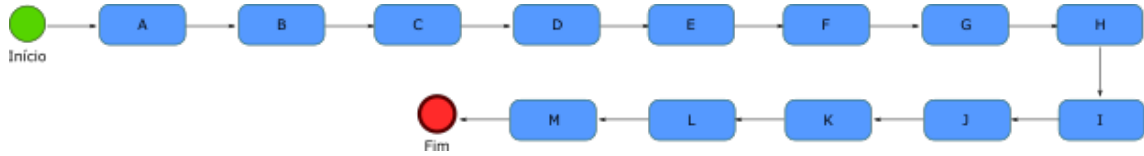


Figura 5.4: Desenho do modelo de Processo 1.3, no padrão de sequência.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,52000	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,92593	1,92593	1,92593	1,96078	1,92593	6,16495	1,92593
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	26,00000	26,00000	26,00000	50,00000	26,00000	299,00000	26,00000
Count Cut Vertices	12,00000	12,00000	12,00000	24,00000	12,00000	0,00000	12,00000
Count Nodes	27,00000	27,00000	27,00000	51,00000	27,00000	97,00000	27,00000
Count Places	14,00000	14,00000	14,00000	26,00000	14,00000	84,00000	14,00000
Count Transitions	13,00000	13,00000	13,00000	25,00000	13,00000	13,00000	13,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,14286	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	1,00000	1,00000	1,00000	1,07692	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,55556	0,55556	0,55556	0,29412	0,55556	0,15464	0,55556
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,92857	1,92857	1,92857	1,96154	1,92857	13,27976	1,92857

Tabela 5.4: Resultados do experimento no padrão de sequência conforme Figura 5.4.

5.2.2 Padrão de Paralelismo

Execução sobre o padrão de paralelismo, primeira opção testada com desvio no modelo. O paralelismo aqui testado está sempre associado com a sincronização dos seus ramos, para evitar mais de um estado final. O modelo inicial consiste em 4 atividades e dois desvios, como mostra a Figura 5.5. Os resultados estão na Tabela 5.5.

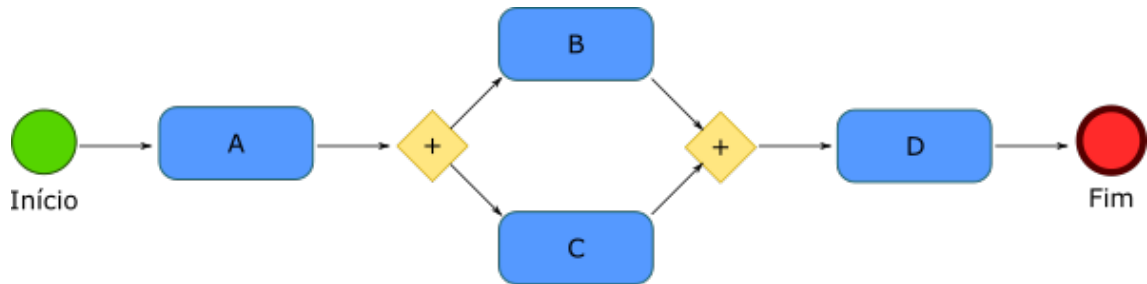


Figura 5.5: Desenho do modelo de Processo 2, no padrão de paralelismo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	0,66667	0,66667	1,00000	1,00000	0,66667
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,15385	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	10,00000	10,00000	14,00000	14,00000	10,00000	27,00000	14,00000
Count Cut Vertices	0,00000	0,00000	2,00000	2,00000	0,00000	0,00000	2,00000
Count Nodes	10,00000	10,00000	14,00000	14,00000	10,00000	13,00000	14,00000
Count Places	6,00000	6,00000	8,00000	8,00000	6,00000	9,00000	8,00000
Count Transitions	4,00000	4,00000	6,00000	6,00000	4,00000	4,00000	6,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,40000	0,40000	1,00000	1,00000	0,40000
Negative Event Generalization	0,54545	0,54545	0,54545	0,54545	0,54545	0,54545	0,54545
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,91667	0,91667	1,25000	1,25000	0,91667	0,91667	1,25000
Simple Structural Appropriateness	0,60000	0,60000	0,42857	0,42857	0,60000	0,46154	0,42857
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,08333	2,08333	2,04167	2,04167	2,08333	4,87500	2,04167

Tabela 5.5: Resultados do experimento no padrão de paralelismo conforme Figura 5.5.

A partir deste caso, houve uma mistura de dois padrões, sendo adicionada uma sequência de atividades entre os dois desvios, em um dos ramos. Ao todo são 5 atividades e 2 desvios, como mostra a Figura 5.6. Os resultados estão na Tabela 5.6.

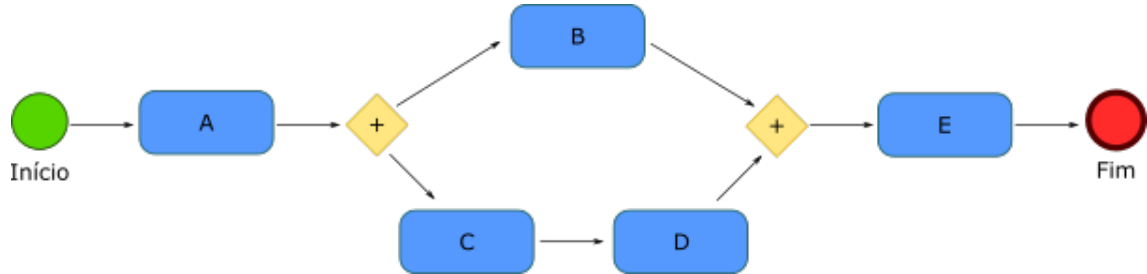


Figura 5.6: Desenho do modelo de Processo 2.1, no padrão de paralelismo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,50000
Advanced Structural Appropriateness	1,00000	1,00000	0,71429	0,62500	1,00000	1,00000	0,71429
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,95050
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,35294	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	12,00000	12,00000	16,00000	18,00000	12,00000	37,00000	16,00000
Count Cut Vertices	1,00000	1,00000	3,00000	4,00000	1,00000	0,00000	2,00000
Count Nodes	12,00000	12,00000	16,00000	18,00000	12,00000	17,00000	16,00000
Count Places	7,00000	7,00000	9,00000	10,00000	7,00000	12,00000	9,00000
Count Transitions	5,00000	5,00000	7,00000	8,00000	5,00000	5,00000	7,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,50000	0,33333	1,00000	1,00000	0,50000
Negative Event Generalization	0,54545	0,54545	0,54545	0,54545	0,54545	0,54545	0,68045
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,95283
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,94173
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,95000	0,95000	1,15000	1,20000	0,95000	0,95000	1,12525
Simple Structural Appropriateness	0,58333	0,58333	0,43750	0,38889	0,58333	0,41176	0,43750
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,05714	2,05714	2,03175	2,02500	2,05714	5,24167	2,03175

Tabela 5.6: Resultados do experimento no padrão de paralelismo conforme Figura 5.6.

Seguindo a linha da extensão anterior, novamente foi acrescido uma tarefa entre os desvios, desta vez, para realizar uma sequência de atividades no ramo que estava com a atividade solitária. A Figura 5.7 mostra como ficou o modelo e a Tabela 5.7 mostra os resultados obtidos.

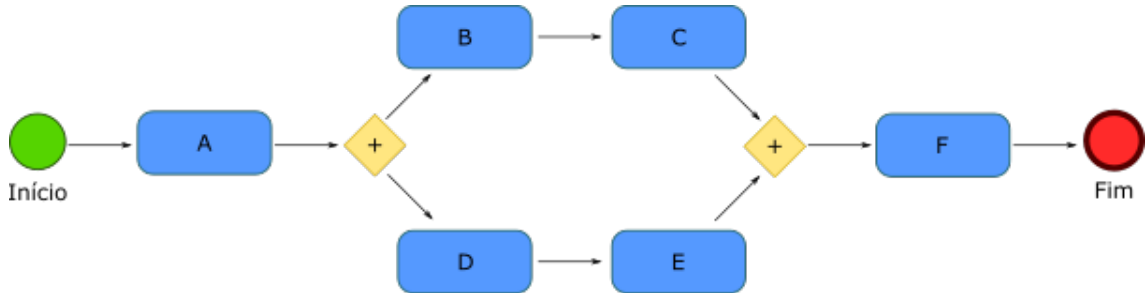


Figura 5.7: Desenho do modelo de Processo 2.2, no padrão de paralelismo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	0,50000	0,50000	0,50000	0,50000	0,66667	1,00000	0,50000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,60000	1,00000	1,00000	0,75000
Alignment Based Fitness	1,00000	1,00000	0,92308	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	0,83333	0,83333	1,00000	0,83333	0,87533	0,91667	0,83333
Alignment Based Probabilistic Generalization	0,99999	0,99999	1,00000	0,99999	0,99999	0,99999	0,99999
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,13333	5,21429	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	14,00000	14,00000	16,00000	22,00000	16,00000	73,00000	18,00000
Count Cut Vertices	0,00000	0,00000	6,00000	2,00000	0,00000	0,00000	2,00000
Count Nodes	14,00000	14,00000	16,00000	22,00000	15,00000	28,00000	18,00000
Count Places	8,00000	8,00000	8,00000	12,00000	9,00000	22,00000	10,00000
Count Transitions	6,00000	6,00000	8,00000	10,00000	6,00000	6,00000	8,00000
ETC Conformance	0,77778	0,77778	0,77699	0,00000	0,82392	0,87500	0,00000
Fitness	1,00000	1,00000	0,85714	0,28571	1,00000	1,00000	0,57143
Negative Event Generalization	0,65693	0,65693	0,44356	0,65693	0,43796	0,43796	0,43796
Negative Event Precision	0,80000	0,80000	0,66888	0,80000	0,80043	0,85714	0,75000
Negative Event Recall	1,00000	1,00000	0,83333	1,00000	1,00000	1,00000	1,00000
One Align Precision	0,77778	0,77778	1,00000	0,78571	0,82392	0,87500	0,81818
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,90000	0,90000	0,92895	1,16667	0,91680	0,93333	1,03333
Simple Structural Appropriateness	0,57143	0,57143	0,50000	0,36364	0,53333	0,28571	0,44444
Succesful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,04167	2,04167	2,00000	2,01667	2,22222	7,74242	2,02500

Tabela 5.7: Resultados do experimento no padrão de paralelismo conforme Figura 5.7.

Voltando à pegada original para este padrão, foram acrescentadas duas atividades antes do estado final, formando uma sequência de atividades posterior ao desvio. O modelo ficou como descrito na Figura 5.8 e os resultados estão apresentados na Tabela 5.8.

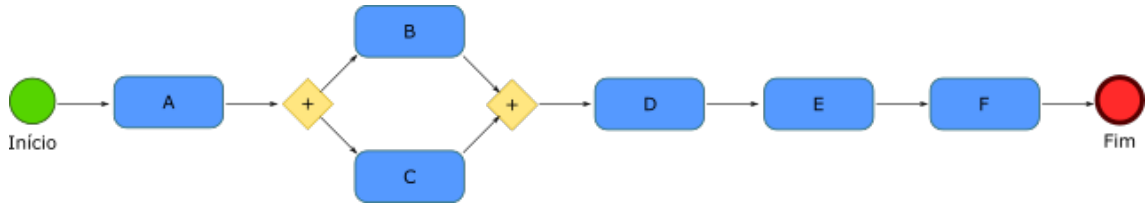


Figura 5.8: Desenho do modelo de Processo 2.3, no padrão de paralelismo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,60000	1,00000	1,00000	0,75000
Alignment Based Fitness	1,00000	1,00000	0,95983	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	0,96282	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,15385	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	14,00000	14,00000	16,00000	22,00000	14,00000	67,00000	18,00000
Count Cut Vertices	2,00000	2,00000	6,00000	6,00000	2,00000	0,00000	4,00000
Count Nodes	14,00000	14,00000	16,00000	22,00000	14,00000	26,00000	18,00000
Count Places	8,00000	8,00000	8,00000	12,00000	8,00000	20,00000	10,00000
Count Transitions	6,00000	6,00000	8,00000	10,00000	6,00000	6,00000	8,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,85714	0,28571	1,00000	1,00000	0,57143
Negative Event Generalization	0,54545	0,54545	0,48855	0,54545	0,54545	0,54545	0,54545
Negative Event Precision	1,00000	1,00000	0,73083	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	0,91967	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,96667	0,96667	0,91767	1,16667	0,96667	0,96667	1,10000
Simple Structural Appropriateness	0,57143	0,57143	0,50000	0,36364	0,57143	0,30769	0,44444
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,04167	2,04167	2,00000	2,01667	2,04167	7,25833	2,02500

Tabela 5.8: Resultados do experimento no padrão de paralelismo conforme Figura 5.8.

Desta vez, replicou-se o modelo inicial deste padrão e acrescentou ao seu final ele novamente. O modelo ficou com 8 atividades e 4 desvios, como descrito na Figura 5.9 e os resultados estão apresentados na Tabela 5.9.

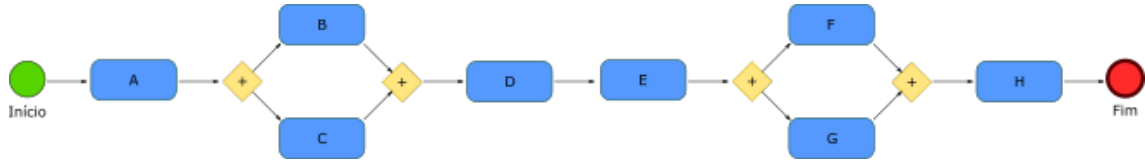


Figura 5.9: Desenho do modelo de Processo 2.4, no padrão de paralelismo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,61538	1,00000	1,00000	0,66667
Alignment Based Fitness	1,00000	1,00000	0,93650	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	0,94366	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,10526	2,10526	2,08696	2,06897	2,10526	5,95833	2,07407
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	20,00000	20,00000	24,00000	30,00000	20,00000	143,00000	28,00000
Count Cut Vertices	1,00000	1,00000	9,00000	6,00000	1,00000	0,00000	5,00000
Count Nodes	19,00000	19,00000	23,00000	29,00000	19,00000	48,00000	27,00000
Count Places	11,00000	11,00000	11,00000	16,00000	11,00000	40,00000	15,00000
Count Transitions	8,00000	8,00000	12,00000	13,00000	8,00000	8,00000	12,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,77778	0,22222	1,00000	1,00000	0,33333
Negative Event Generalization	0,63636	0,63636	0,50567	0,63636	0,63636	0,63636	0,63636
Negative Event Precision	1,00000	1,00000	0,60906	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	0,87300	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,96429	0,96429	0,91622	1,12500	0,96429	0,96429	1,10714
Simple Structural Appropriateness	0,52632	0,52632	0,43478	0,34483	0,52632	0,20833	0,37037
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,15909	2,15909	2,09091	2,09135	2,15909	10,72500	2,10000

Tabela 5.9: Resultados do experimento no padrão de paralelismo conforme Figura 5.9.

5.2.3 Padrão de Desvio Exclusivo

O desvio exclusivo aqui testado está sempre associado com a convergência dos seus ramos, para evitar mais de um estado final. O primeiro modelo deste padrão tem 4 atividades e 2 desvios, como mostra a Figura 5.10. Os resultados estão descritos na Tabela 5.10.

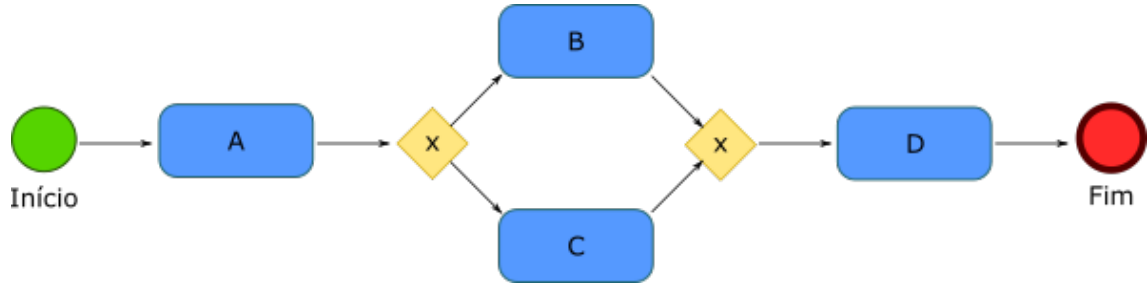


Figura 5.10: Desenho do modelo de Processo 3, no padrão de desvio exclusivo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	3,25000	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	8,00000	8,00000	8,00000	16,00000	8,00000	13,00000	8,00000
Count Cut Vertices	2,00000	2,00000	2,00000	2,00000	2,00000	0,00000	2,00000
Count Nodes	8,00000	8,00000	8,00000	16,00000	8,00000	8,00000	8,00000
Count Places	4,00000	4,00000	4,00000	8,00000	4,00000	4,00000	4,00000
Count Transitions	4,00000	4,00000	4,00000	8,00000	4,00000	4,00000	4,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,88889	0,88889	0,88889	1,22222	0,88889	0,88889	0,88889
Simple Structural Appropriateness	0,75000	0,75000	0,75000	0,37500	0,75000	0,75000	0,75000
Succesful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	3,25000	2,00000

Tabela 5.10: Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.10.

Em continuidade neste padrão, o primeiro acréscimo foi de duas atividades em cada ramo do desvio, fazendo dentro deles uma sequência de atividades, ficando como mostra a Figura 5.11. Os resultados estão descritos na Tabela 5.11

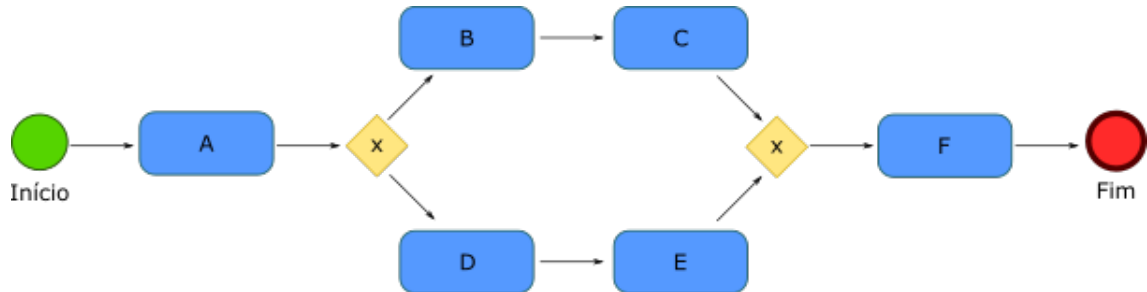


Figura 5.11: Desenho do modelo de Processo 3.1, no padrão de desvio exclusivo.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	0,83333	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	0,93925	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,88889	2,00000
Behavioral Profile Conformance	1,00000	1,00000	0,90280	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	12,00000	12,00000	12,00000	24,00000	12,00000	44,00000	12,00000
Count Cut Vertices	2,00000	2,00000	3,00000	2,00000	2,00000	0,00000	2,00000
Count Nodes	12,00000	12,00000	12,00000	24,00000	12,00000	18,00000	12,00000
Count Places	6,00000	6,00000	6,00000	12,00000	6,00000	12,00000	6,00000
Count Transitions	6,00000	6,00000	6,00000	12,00000	6,00000	6,00000	6,00000
ETC Conformance	1,00000	1,00000	0,85488	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	0,90280	0,40000	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	0,48600	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	0,70280	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	0,87850	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,95000	0,95000	0,92570	1,15000	0,95000	0,95000	0,95000
Simple Structural Appropriateness	0,66667	0,66667	0,66667	0,33333	0,66667	0,44444	0,66667
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,50000	2,00000

Tabela 5.11: Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.11.

Voltando ao modelo original deste padrão, foi feito um acréscimo de duas atividades anteriores ao desvio, fazendo uma sequência de atividades, ficando como mostra a Figura 5.12. Os resultados estão descritos na Tabela 5.12

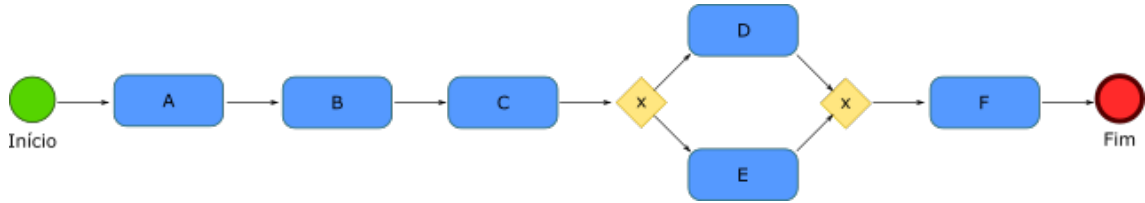


Figura 5.12: *Desenho do modelo de Processo 3.2, no padrão de desvio exclusivo.*

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,00000	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	12,00000	12,00000	12,00000	24,00000	12,00000	45,00000	12,00000
Count Cut Vertices	4,00000	4,00000	4,00000	6,00000	4,00000	0,00000	4,00000
Count Nodes	12,00000	12,00000	12,00000	24,00000	12,00000	18,00000	12,00000
Count Places	6,00000	6,00000	6,00000	12,00000	6,00000	12,00000	6,00000
Count Transitions	6,00000	6,00000	6,00000	12,00000	6,00000	6,00000	6,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,33333	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,96000	0,96000	0,96000	1,16000	0,96000	0,96000	0,96000
Simple Structural Appropriateness	0,66667	0,66667	0,66667	0,33333	0,66667	0,44444	0,66667
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,62500	2,00000

Tabela 5.12: *Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.12.*

Desta vez, replicou-se o processo inicial deste padrão e replicou-o no final, ficando como mostra a Figura 5.13. Os resultados estão descritos na Tabela 5.13

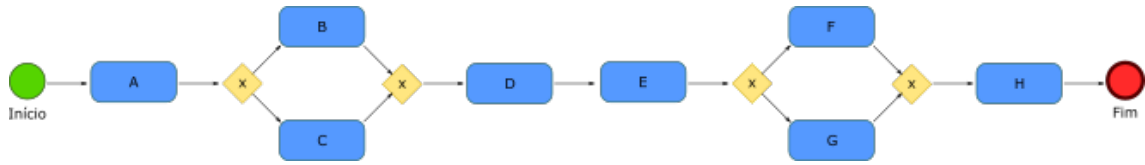


Figura 5.13: *Desenho do modelo de Processo 3.3, no padrão de desvio exclusivo.*

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,47059	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	0,83617	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,13333	2,13333	1,84615	2,06061	2,13333	6,07407	2,13333
Behavioral Profile Conformance	1,00000	1,00000	0,83617	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	16,00000	16,00000	12,00000	34,00000	16,00000	82,00000	16,00000
Count Cut Vertices	5,00000	5,00000	5,00000	6,00000	5,00000	0,00000	5,00000
Count Nodes	15,00000	15,00000	13,00000	33,00000	15,00000	27,00000	15,00000
Count Places	7,00000	7,00000	7,00000	16,00000	7,00000	19,00000	7,00000
Count Transitions	8,00000	8,00000	6,00000	17,00000	8,00000	8,00000	8,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	0,83663	0,28571	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	0,64585	1,00000	1,00000	1,00000	1,00000
Negative Event Recall	1,00000	1,00000	0,67233	1,00000	1,00000	1,00000	1,00000
One Align Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,95238	0,95238	0,94602	1,11905	0,95238	0,95238	0,95238
Simple Structural Appropriateness	0,66667	0,66667	0,61538	0,30303	0,66667	0,37037	0,66667
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,14286	2,14286	1,85714	2,06250	2,14286	7,28289	2,14286

Tabela 5.13: *Resultados do experimento no padrão de desvio exclusivo conforme Figura 5.13.*

5.2.4 Padrão de Laço

O laço aqui testado está associado a desvios, não havendo laço de uma só atividade. O primeiro modelo deste padrão tem 4 atividades e 2 desvios, como mostra a Figura 5.14. Os resultados estão descritos na Tabela 5.14.

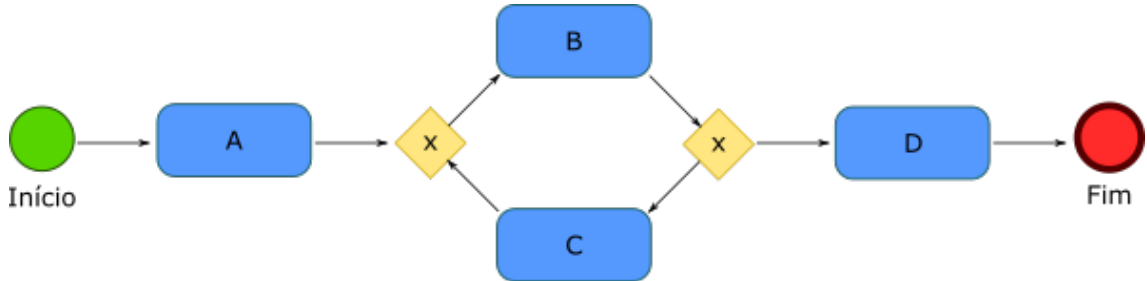


Figura 5.14: *Desenho do modelo de Processo 4, no padrão de laço.*

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	0,33333	0,69444	0,69444	0,69444	0,69444	0,69444	0,69444
Advanced Structural Appropriateness	1,00000	1,00000	0,80000	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	-1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	0,00000	0,99276	0,99276	0,99276	0,99276	0,99276	0,99276
Alignment Based Probabilistic Generalization	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,50000	2,00000	2,00000	2,00000	2,00000	4,20000	2,00000
Behavioral Profile Conformance	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	6,00000	8,00000	10,00000	16,00000	8,00000	21,00000	8,00000
Count Cut Vertices	2,00000	2,00000	3,00000	4,00000	2,00000	0,00000	2,00000
Count Nodes	8,00000	8,00000	10,00000	16,00000	8,00000	10,00000	8,00000
Count Places	4,00000	4,00000	5,00000	8,00000	4,00000	6,00000	4,00000
Count Transitions	4,00000	4,00000	5,00000	8,00000	4,00000	4,00000	4,00000
ETC Conformance	0,53397	0,98964	0,00000	0,00000	0,98964	0,98964	0,98964
Fitness	0,80792	1,00000	0,83057	0,33887	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	0,50176	0,99567	0,99567	0,99567	0,99567	0,99567	0,99567
Negative Event Recall	0,80600	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	0,60462	0,98964	0,99096	0,99340	0,98964	0,98964	0,98964
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,66667	0,87459	1,00000	1,25082	0,87459	0,87459	0,87459
Simple Structural Appropriateness	0,75000	0,75000	0,60000	0,37500	0,75000	0,60000	0,75000
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,50000	2,00000	2,00000	2,00000	2,00000	4,37500	2,00000

Tabela 5.14: *Resultados do experimento no padrão de laço conforme Figura 5.14.*

Houve agora um acréscimo de uma atividade na primeira parte do laço, como mostra a Figura 5.15. Os resultados estão descritos na Tabela 5.15.

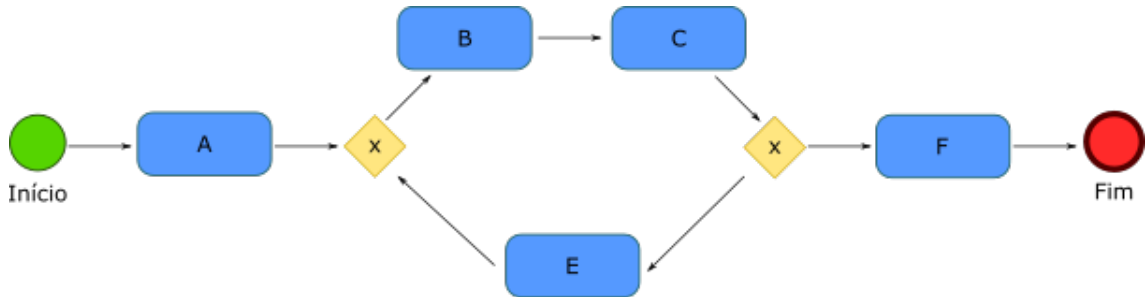


Figura 5.15: Desenho do modelo de Processo 4.1, no padrão de laço.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	0,75214	0,75214	0,75214	0,75214	0,75214	0,75214	0,75214
Advanced Structural Appropriateness	1,00000	1,00000	0,85714	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	0,99596	0,99596	1,00000	0,99596	0,99596	0,99596	0,99596
Alignment Based Probabilistic Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,30769	2,00000
Behavioral Profile Conformance	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	10,00000	10,00000	14,00000	20,00000	10,00000	28,00000	10,00000
Count Cut Vertices	2,00000	2,00000	4,00000	4,00000	2,00000	0,00000	2,00000
Count Nodes	10,00000	10,00000	14,00000	20,00000	10,00000	13,00000	10,00000
Count Places	5,00000	5,00000	7,00000	10,00000	5,00000	8,00000	5,00000
Count Transitions	5,00000	5,00000	7,00000	10,00000	5,00000	5,00000	5,00000
ETC Conformance	0,99371	0,99371	0,00000	0,00000	0,99371	0,99371	0,99371
Fitness	1,00000	1,00000	0,86993	0,26015	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	0,52381	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	0,99785	0,99785	0,99747	0,99785	0,99785	0,99785	0,99785
Negative Event Recall	1,00000	1,00000	0,85048	1,00000	1,00000	1,00000	1,00000
One Align Precision	0,99371	0,99371	0,99437	0,99622	0,99371	0,99371	0,99371
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,93197	0,93197	1,00000	1,20409	0,93197	0,93197	0,93197
Simple Structural Appropriateness	0,70000	0,70000	0,50000	0,35000	0,70000	0,53846	0,70000
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,55000	2,00000

Tabela 5.15: Resultados do experimento no padrão de laço conforme Figura 5.15.

Voltando ao modelo original neste padrão, houve um acréscimo de uma atividade na antes do estado final, como mostra a Figura 5.16. Os resultados estão descritos na Tabela 5.16.

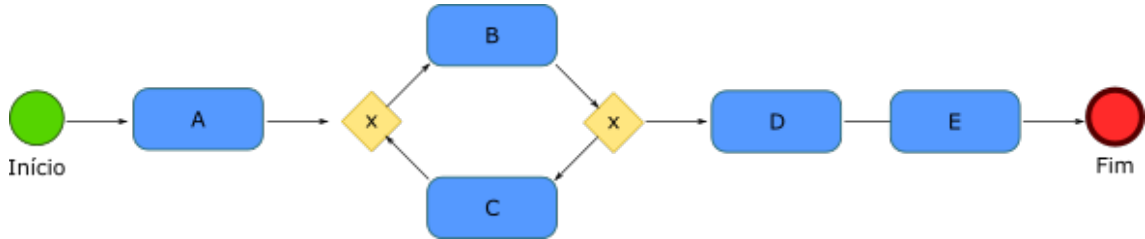


Figura 5.16: *Desenho do modelo de Processo 4.2, no padrão de laço.*

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	0,37500	0,56731	0,56731	0,56731	0,56731	0,56731	0,56731
Advanced Structural Appropriateness	1,00000	1,00000	0,83333	0,50000	1,00000	1,00000	1,00000
Alignment Based Fitness	-1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	0,00000	0,99444	0,99444	0,99444	0,99444	0,99444	0,99444
Alignment Based Probabilistic Generalization	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	2,00000	2,00000	2,00000	2,00000	4,80000	2,00000
Behavioral Profile Conformance	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	8,00000	10,00000	12,00000	20,00000	10,00000	36,00000	10,00000
Count Cut Vertices	3,00000	3,00000	4,00000	6,00000	3,00000	0,00000	3,00000
Count Nodes	10,00000	10,00000	12,00000	20,00000	10,00000	15,00000	10,00000
Count Places	5,00000	5,00000	6,00000	10,00000	5,00000	10,00000	5,00000
Count Transitions	5,00000	5,00000	6,00000	10,00000	5,00000	5,00000	5,00000
ETC Conformance	0,52774	0,99167	0,00000	0,00000	0,99167	0,99167	0,99167
Fitness	0,83153	1,00000	0,85767	0,28466	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	0,47341	0,99668	0,99668	0,99668	0,99668	0,99668	0,99668
Negative Event Recall	0,83190	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
One Align Precision	0,59541	0,99167	0,99259	0,99487	0,99167	0,99167	0,99167
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,73194	0,92243	0,95257	1,20257	0,92243	0,92243	0,92243
Simple Structural Appropriateness	0,70000	0,70000	0,58333	0,35000	0,70000	0,46667	0,70000
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,60000	2,00000	2,00000	2,00000	2,00000	5,40000	2,00000

Tabela 5.16: *Resultados do experimento no padrão de laço conforme Figura 5.16.*

Desta vez, replicou-se o modelo inicial deste padrão e acrescentou ao seu final ele novamente, ficando como mostra a Figura 5.17. Os resultados estão descritos na Tabela 5.17.

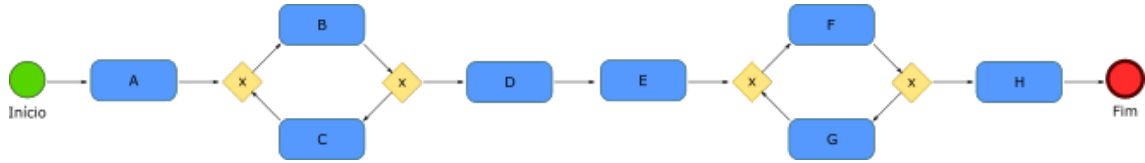


Figura 5.17: Desenho do modelo de Processo 4.3, no padrão de laço.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Behavioural Appropriateness	0,32143	0,59477	0,59477	0,59477	0,59477	0,59477	0,59477
Advanced Structural Appropriateness	1,00000	1,00000	0,75000	0,47059	1,00000	1,00000	1,00000
Alignment Based Fitness	-1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Alignment Based Precision	0,00000	0,99394	0,99642	0,99394	0,99394	0,99394	0,99394
Alignment Based Probabilistic Generalization	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	2,13333	2,08696	2,06061	2,13333	5,85714	2,13333
Behavioral Profile Conformance	0,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Best Align Precision	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Count Arcs	12,00000	16,00000	24,00000	34,00000	16,00000	82,00000	16,00000
Count Cut Vertices	5,00000	5,00000	8,00000	10,00000	5,00000	0,00000	5,00000
Count Nodes	15,00000	15,00000	23,00000	33,00000	15,00000	28,00000	15,00000
Count Places	7,00000	7,00000	11,00000	16,00000	7,00000	20,00000	7,00000
Count Transitions	8,00000	8,00000	12,00000	17,00000	8,00000	8,00000	8,00000
ETC Conformance	0,37109	0,99132	0,00000	0,00000	0,99132	0,99132	0,99132
Fitness	0,79329	1,00000	0,64012	0,18783	1,00000	1,00000	1,00000
Negative Event Generalization	1,00000	1,00000	0,77959	1,00000	1,00000	1,00000	1,00000
Negative Event Precision	0,28525	0,99493	0,63211	0,99493	0,99493	0,99493	0,99493
Negative Event Recall	0,81095	1,00000	0,89635	1,00000	1,00000	1,00000	1,00000
One Align Precision	0,52986	0,99132	0,99324	0,99471	0,99132	0,99132	0,99132
Proper Completion	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Simple Behavioural Appropriateness	0,68681	0,94505	0,95903	1,12638	0,94505	0,94505	0,94505
Simple Structural Appropriateness	0,66667	0,66667	0,43478	0,30303	0,66667	0,35714	0,66667
Successful Execution	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000	0,00000
Weighed P/T Average Arc Degree	1,60714	2,14286	2,09091	2,06250	2,14286	7,17500	2,14286

Tabela 5.17: Resultados do experimento no padrão de laço conforme Figura 5.17.

Análises

Neste capítulo são apresentados dados decorrentes da análise dos experimentos realizados. Ao todo foram geradas 119 linhas de registros, dos 17 experimentos, que contém as 25 métricas aplicadas para cada um dos 6 modelos gerados pelos algoritmos de Descoberta de Processo mais o modelo original.

6.1 Seleção de Atributos

Para a Seleção de Atributos, utilizando o WEKA, a aplicação do componente **CfsSubsetEval** com **BestFirst** selecionaram as seguintes 8 métricas como relevantes: *Advanced Structural Appropriateness, Average Node Arc Degree, Count Arcs, ETC Conformance, Fitness, Negative Event Precision, Simple Structural Appropriateness, Weighed PT Average Arc Degree*.

Observa-se dessas métricas que 3 avaliam Aptidão, 3 avaliam Precisão, 2 avaliam Simplicidade e nenhuma avalia Generalização. Com elas, é possível reduzir a quantidade de dados a analisar, o que facilita a comparação manual. Nas próximas seções, seguem dados filtrados e comparados.

6.2 Padrão de Sequência de Atividades

Para este padrão de sequência de atividades, notou-se uma grande equivalência de métricas do modelo original com os algoritmos Alpha, Alpha#, ETMd e ILP Miner, como destacado nas Tabelas 6.1 e 6.4. Das Tabela 6.2 e 6.3, o ILP Miner não conseguiu reproduzir todas as métricas. Em todos os casos, o Heuristics Miner e o ILP-Based não conseguiram bons resultados.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,66667	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	1,60000	1,60000	1,71429	1,60000	1,50000	1,60000
Count Arcs	4,00000	4,00000	4,00000	6,00000	4,00000	3,00000	4,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,66667	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,80000	0,80000	0,80000	0,57143	0,80000	1,00000	0,80000
Weighed P/T Average Arc Degree	1,66667	1,66667	1,66667	1,75000	1,66667	1,50000	1,66667

Tabela 6.1: Métricas relevantes do Experimento 5.1.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,55556	0,75000	1,00000	1,00000
Average Node Arc Degree	1,81818	1,81818	1,81818	1,89474	1,88235	4,11111	1,81818
Count Arcs	10,00000	10,00000	10,00000	18,00000	16,00000	37,00000	10,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,33333	0,83333	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,63636	0,63636	0,63636	0,36842	0,41176	0,38889	0,63636
Weighed P/T Average Arc Degree	1,83333	1,83333	1,83333	1,90000	1,88889	5,12308	1,83333

Tabela 6.2: Métricas relevantes do Experimento 5.2.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,53333	1,00000	1,00000	1,00000
Average Node Arc Degree	1,88235	1,88235	1,88235	1,93548	1,87500	5,28205	1,88235
Count Arcs	16,00000	16,00000	16,00000	30,00000	15,00000	103,00000	16,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,22222	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,58824	0,58824	0,58824	0,32258	0,62500	0,25641	0,58824
Weighed P/T Average Arc Degree	1,88889	1,88889	1,88889	1,93750	1,87500	8,09879	1,88889

Tabela 6.3: Métricas relevantes do Experimento 5.3.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,52000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,92593	1,92593	1,92593	1,96078	1,92593	6,16495	1,92593
Count Arcs	26,00000	26,00000	26,00000	50,00000	26,00000	299,00000	26,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,14286	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,55556	0,55556	0,55556	0,29412	0,55556	0,15464	0,55556
Weighed P/T Average Arc Degree	1,92857	1,92857	1,92857	1,96154	1,92857	13,27976	1,92857

Tabela 6.4: Métricas relevantes do Experimento 5.4.

6.3 Padrão de Paralelismo

Para o padrão de paralelismo, houve um bom resultado dos algoritmos ETMd e Heuristics Miner, mas apenas para o caso apresentado na Tabela 6.5. O ETMd novamente se apresenta razoável, como mostra a Tabela 6.6. Mas, em todos os demais casos, nenhum algoritmo conseguiu equivalência de métricas com o modelo Original.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,66667	0,66667	1,00000	1,00000	0,66667
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,15385	2,00000
Count Arcs	10,00000	10,00000	14,00000	14,00000	10,00000	27,00000	14,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,40000	0,40000	1,00000	1,00000	0,40000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,60000	0,60000	0,42857	0,42857	0,60000	0,46154	0,42857
Weighed P/T Average Arc Degree	2,08333	2,08333	2,04167	2,04167	2,08333	4,87500	2,04167

Tabela 6.5: Métricas relevantes do Experimento 5.5.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,71429	0,62500	1,00000	1,00000	0,71429
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,35294	2,00000
Count Arcs	12,00000	12,00000	16,00000	18,00000	12,00000	37,00000	16,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,50000	0,33333	1,00000	1,00000	0,50000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	0,95283
Simple Structural Appropriateness	0,58333	0,58333	0,43750	0,38889	0,58333	0,41176	0,43750
Weighed P/T Average Arc Degree	2,05714	2,05714	2,03175	2,02500	2,05714	5,24167	2,03175

Tabela 6.6: Métricas relevantes do Experimento 5.6.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,60000	1,00000	1,00000	0,75000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,13333	5,21429	2,00000
Count Arcs	14,00000	14,00000	16,00000	22,00000	16,00000	73,00000	18,00000
ETC Conformance	0,77778	0,77778	0,77699	0,00000	0,82392	0,87500	0,00000
Fitness	1,00000	1,00000	0,85714	0,28571	1,00000	1,00000	0,57143
Negative Event Precision	0,80000	0,80000	0,66888	0,80000	0,80043	0,85714	0,75000
Simple Structural Appropriateness	0,57143	0,57143	0,50000	0,36364	0,53333	0,28571	0,44444
Weighed P/T Average Arc Degree	2,04167	2,04167	2,00000	2,01667	2,22222	7,74242	2,02500

Tabela 6.7: Métricas relevantes do Experimento 5.7.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,60000	1,00000	1,00000	0,75000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,15385	2,00000
Count Arcs	14,00000	14,00000	16,00000	22,00000	14,00000	67,00000	18,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,85714	0,28571	1,00000	1,00000	0,57143
Negative Event Precision	1,00000	1,00000	0,73083	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,57143	0,57143	0,50000	0,36364	0,57143	0,30769	0,44444
Weighed P/T Average Arc Degree	2,04167	2,04167	2,00000	2,01667	2,04167	7,25833	2,02500

Tabela 6.8: Métricas relevantes do Experimento 5.8.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,61538	1,00000	1,00000	0,66667
Average Node Arc Degree	2,10526	2,10526	2,08696	2,06897	2,10526	5,95833	2,07407
Count Arcs	20,00000	20,00000	24,00000	30,00000	20,00000	143,00000	28,00000
ETC Conformance	1,00000	1,00000	0,00000	0,00000	1,00000	1,00000	0,00000
Fitness	1,00000	1,00000	0,77778	0,22222	1,00000	1,00000	0,33333
Negative Event Precision	1,00000	1,00000	0,60906	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,52632	0,52632	0,43478	0,34483	0,52632	0,20833	0,37037
Weighed P/T Average Arc Degree	2,15909	2,15909	2,09091	2,09135	2,15909	10,72500	2,10000

Tabela 6.9: Métricas relevantes do Experimento 5.9.

6.4 Padrão de Desvio Exclusivo

Para o padrão de desvio exclusivo, como mostram as Tabelas 6.10, 6.11, 6.12 e 6.13 os algoritmos Alpha, Alpha# e ILP apresentaram resultados equivalentes ao modelo Original. O ETMd conseguiu equivalência ao Original nos resultados apresentados pelas Tabelas 6.10 e 6.12. O Heuristics Miner e o ILP-Based não conseguiram bons resultados.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	3,25000	2,00000
Count Arcs	8,00000	8,00000	8,00000	16,00000	8,00000	13,00000	8,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,75000	0,75000	0,75000	0,37500	0,75000	0,75000	0,75000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	3,25000	2,00000

Tabela 6.10: Métricas relevantes do Experimento 5.10.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,88889	2,00000
Count Arcs	12,00000	12,00000	12,00000	24,00000	12,00000	44,00000	12,00000
ETC Conformance	1,00000	1,00000	0,85488	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	0,90280	0,40000	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	0,70280	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,66667	0,66667	0,66667	0,33333	0,66667	0,44444	0,66667
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,50000	2,00000

Tabela 6.11: Métricas relevantes do Experimento 5.11.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,00000	2,00000
Count Arcs	12,00000	12,00000	12,00000	24,00000	12,00000	45,00000	12,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	1,00000	0,33333	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,66667	0,66667	0,66667	0,33333	0,66667	0,44444	0,66667
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	5,62500	2,00000

Tabela 6.12: Métricas relevantes do Experimento 5.12.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	1,00000	0,47059	1,00000	1,00000	1,00000
Average Node Arc Degree	2,13333	2,13333	1,84615	2,06061	2,13333	6,07407	2,13333
Count Arcs	16,00000	16,00000	12,00000	34,00000	16,00000	82,00000	16,00000
ETC Conformance	1,00000	1,00000	1,00000	0,00000	1,00000	1,00000	1,00000
Fitness	1,00000	1,00000	0,83663	0,28571	1,00000	1,00000	1,00000
Negative Event Precision	1,00000	1,00000	0,64585	1,00000	1,00000	1,00000	1,00000
Simple Structural Appropriateness	0,66667	0,66667	0,61538	0,30303	0,66667	0,37037	0,66667
Weighed P/T Average Arc Degree	2,14286	2,14286	1,85714	2,06250	2,14286	7,28289	2,14286

Tabela 6.13: Métricas relevantes do Experimento 5.13.

6.5 Padrão de Laço

Para o padrão de laço, os algoritmos Alpha# e ILP conseguiram equivalência com o Original, como mostram as Tabelas 6.14, 6.15, 6.16 e 6.17. Fora eles, apenas o algoritmo Alpha conseguiu equivalência nas métricas para um experimento, como apresenta a Tabela 6.15. Os demais algoritmos, ETMd, Heuristics Miner, ILP-Based, não conseguiram bons resultados.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,80000	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,50000	2,00000	2,00000	2,00000	2,00000	4,20000	2,00000
Count Arcs	6,00000	8,00000	10,00000	16,00000	8,00000	21,00000	8,00000
ETC Conformance	0,53397	0,98964	0,00000	0,00000	0,98964	0,98964	0,98964
Fitness	0,80792	1,00000	0,83057	0,33887	1,00000	1,00000	1,00000
Negative Event Precision	0,50176	0,99567	0,99567	0,99567	0,99567	0,99567	0,99567
Simple Structural Appropriateness	0,75000	0,75000	0,60000	0,37500	0,75000	0,60000	0,75000
Weighed P/T Average Arc Degree	1,50000	2,00000	2,00000	2,00000	2,00000	4,37500	2,00000

Tabela 6.14: Métricas relevantes do Experimento 5.14.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,85714	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,30769	2,00000
Count Arcs	10,00000	10,00000	14,00000	20,00000	10,00000	28,00000	10,00000
ETC Conformance	0,99371	0,99371	0,00000	0,00000	0,99371	0,99371	0,99371
Fitness	1,00000	1,00000	0,86993	0,26015	1,00000	1,00000	1,00000
Negative Event Precision	0,99785	0,99785	0,99747	0,99785	0,99785	0,99785	0,99785
Simple Structural Appropriateness	0,70000	0,70000	0,50000	0,35000	0,70000	0,53846	0,70000
Weighed P/T Average Arc Degree	2,00000	2,00000	2,00000	2,00000	2,00000	4,55000	2,00000

Tabela 6.15: Métricas relevantes do Experimento 5.15.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,83333	0,50000	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	2,00000	2,00000	2,00000	2,00000	4,80000	2,00000
Count Arcs	8,00000	10,00000	12,00000	20,00000	10,00000	36,00000	10,00000
ETC Conformance	0,52774	0,99167	0,00000	0,00000	0,99167	0,99167	0,99167
Fitness	0,83153	1,00000	0,85767	0,28466	1,00000	1,00000	1,00000
Negative Event Precision	0,47341	0,99668	0,99668	0,99668	0,99668	0,99668	0,99668
Simple Structural Appropriateness	0,70000	0,70000	0,58333	0,35000	0,70000	0,46667	0,70000
Weighed P/T Average Arc Degree	1,60000	2,00000	2,00000	2,00000	2,00000	5,40000	2,00000

Tabela 6.16: Métricas relevantes do Experimento 5.16.

	Alpha	Alpha#	ETMd	Heuristics Miner	ILP	ILP-Based	Original
Advanced Structural Appropriateness	1,00000	1,00000	0,75000	0,47059	1,00000	1,00000	1,00000
Average Node Arc Degree	1,60000	2,13333	2,08696	2,06061	2,13333	5,85714	2,13333
Count Arcs	12,00000	16,00000	24,00000	34,00000	16,00000	82,00000	16,00000
ETC Conformance	0,37109	0,99132	0,00000	0,00000	0,99132	0,99132	0,99132
Fitness	0,79329	1,00000	0,64012	0,18783	1,00000	1,00000	1,00000
Negative Event Precision	0,28525	0,99493	0,63211	0,99493	0,99493	0,99493	0,99493
Simple Structural Appropriateness	0,66667	0,66667	0,43478	0,30303	0,66667	0,35714	0,66667
Weighed P/T Average Arc Degree	1,60714	2,14286	2,09091	2,06250	2,14286	7,17500	2,14286

Tabela 6.17: Métricas relevantes do Experimento 5.17.

6.6 Resumo

Analisando as equivalências, compilam-se os resultados obtidos anteriormente no seguinte resumo, Tabela 6.18.

Algoritmos	Sequência (4)	Paralelismo (5)	Desvio Exclusivo (4)	Laço (4)	Total (17)
Alpha#	4	0	4	4	12
ILP	2	0	4	4	10
Alpha	4	0	4	1	9
ETMd	4	1	2	0	7
Heuristics Miner	0	1	0	0	1
ILP-Based	0	0	0	0	0

Tabela 6.18: *Resumo dos algoritmos que conseguiram equivalência com modelo original por padrão.*

Percebe-se dessa compilação que: o melhor algoritmo para esses experimentos foi o Alpha#, que o padrão de paralelismo se mostrou um desafio para os algoritmos, que o Heuristics Miner e o ILP-Based tiveram baixo desempenho.

6.7 Agrupamentos

Outra análise realizada ao classificar utilizando o WEKA, foram obtidos com o Agrupamento realizado pelo componente **EM**, seguem dados na Tabela 6.19. O resultado apresenta a descoberta de dois agrupamentos, que separam o ILP-Based dos demais algoritmos. Ressalta-se dessa tabela o Heuristics Miner, que não foi bem em equivalência com o Original, no Agrupamento 1.

Agrupamento 0	Agrupamento 1	
0	17	Heuristics Miner
1	16	Original
1	16	Alpha#
1	16	ILP
2	15	ETMd
4	13	Alpha
16	1	ILP-Based

Tabela 6.19: *Aplicação do componente EM para classificar Algoritmos.*

Ao realizar substituição dos Algoritmos por identificação do padrão, o resultado na Tabela 6.20, mostra que os padrões mais se assemelharam ao de sequência, ficando uma média de 75% da classificação no mesmo agrupamento, com maior proporção desse padrão.

Agrupamento 0	Agrupamento 1	
3	25	Sequência
5	23	Desvio Exclusivo
7	21	Lço
10	25	Paralelismo

Tabela 6.20: Aplicação do componente *EM* para classificar Padrão de Fluxo de Trabalho.

Ao realizar outra substituição dos Algoritmos por sua família, como mostra a Tabela 6.21, vemos dois agrupamentos com o primeiro sendo dominado pela família de Duas Fases, família do algoritmo ILP-Based.

Agrupamento 0	Agrupamento 1	
2	15	Inteligência Computacional
5	46	Direta
17	17	Duas Fases

Tabela 6.21: Aplicação do componente *EM* para classificar Família.

Essas análises realizadas com o complemento de Agrupamento do WEKA revelam a necessidade de apenas dois agrupamentos.

6.8 Geral

Apesar de seminal, este tabalho está em contraste com van der Aalst [24], no quesito de classificação, pois foram necessários apenas dois agrupamentos para separar os algoritmos, o que revela que a separação em abordagem computacional, em quatro grupos, não tem relação com a competência dos algoritmos.

Os resultados experimentais também revelam a dificuldade em se obter modelos equivalentes com o esperado devido ao padrão de paralelismo, e até mesmo o de laço, que apresentou uma queda em relação ao de sequência e desvio exclusivo, o que está de acordo com van der Aalst[24], onde havia sido apontado esse desafio.

Em relação ao modelo de Rozinat et al [23], o ambiente proposto foi capaz de reproduzir o fluxo do mesmo, simulando a iteração de dados, com a execução de algoritmos sobre os registros e a comparação dos modelos de processo gerados com métricas de qualidade.

Em relação aos modelos de Wang et al [36] e Ribeiro et al [22], o ambiente proposto também foi capaz de classificar os algoritmos com base na aplicação das métricas nos modelos de processo gerados, tendo resultados por padrão de fluxo de trabalho e geral.

Conclusão e Trabalhos Futuros

Neste capítulo estão as conclusões decorrentes deste trabalho e sugestões para trabalhos futuros.

7.1 Conclusão

O objetivo principal deste trabalho foi alcançado, pois, como se observou dos resultados obtidos, na Seção 6, o ambiente proposto foi capaz de classificar, avaliar e comparar os algoritmos.

A primeira observação decorrente deste trabalho é sobre os padrões da Mineração de Processos. A evolução dos algoritmos mostra padrões de entrada bem definidos, mas, para a saída, ainda há um porém, que tem como restrição a conversão para Rede de Petri quando não é este o seu formato original de saída, para uso em demais fases da Mineração de Processos. Essa é, definitivamente, uma barreira que deve ser encarada na área.

Outra importante observação, ocorrida nos experimentos, é a necessidade de mais detalhes presentes nos registros dos eventos. Pois existem casos em que apenas com informações adicionais é possível determinar corretamente um paralelismo, por exemplo.

Os resultados alcançados trouxeram respostas para os objetivos levantados na Seção 1.1:

1. Ao fixar os padrões de entrada, foi possível constatar dois agrupamentos, como está destacado na Seção 6.7, em que o ILP-Based ficou separado dos demais algoritmos;
2. Com a análise de Seleção de Atributos, realizada na Seção 6.1, observou-se que de 24 métricas, apenas 8 eram suficientes para classificar os algoritmos;
3. Como mostra a Tabela 6.19, foi possível com o ambiente utilizado, descrito na Seção 4.1, classificar os algoritmos por relevância para cada padrão e de maneira geral.

O ambiente utilizado provou-se promissor no sentido de ser um padrão de avaliação e classificação de algoritmos, requerendo para isso seu uso e o mantimento dos resultados em um repositório público e disponível.

7.2 Trabalhos Futuros

Apesar do sucesso experimental obtido, o ambiente utilizado pode ser melhorado se houver um maior número de padrões de fluxo de trabalho, métricas e algoritmos. Assim, sugere-se para futuros trabalhos: 1) desenvolver modelos para outros padrões, além dos quatro utilizados; 2) utilizar mais métricas e avaliar suas relevâncias; 3) aplicar nesse ambiente mais algoritmos.

Outra observação é sobre as ferramentas utilizadas, que apesar de serem suficientes e bastante úteis, nem sempre apresentam interface amigável ou não estão totalmente funcionais, de modo que, pensar numa automação do ambiente utilizado não parece um trabalho trivial, mas, ainda sim, é recomendado como trabalho futuro.

Referências Bibliográficas

- [1] ALPAYDIN, E. **Introduction to Machine Learning Second Edition**. The MIT Press, London, England, 2nd edition, 2010.
- [2] BECKER, M.; LAUE, R. **A comparative survey of business process similarity measures**. *Computers in Industry*, 63(2):148–167, 2012.
- [3] BERGENTHUM, R.; MAUSER, S. **Synthesis of petri nets from infinite partial languages with viptool**. *CEUR Workshop Proceedings*, 380:81–86, 2008.
- [4] BLUM, F. R. **Metrics in process discovery**. Technical Report 2015-6, 2015.
- [5] BUIJS, J. C. A. M.; VAN DONGEN, B. F.; VAN DER AALST, W. M. P. **A genetic algorithm for discovering process trees**. *2012 IEEE Congress on Evolutionary Computation, CEC 2012*, p. 10–15, 2012.
- [6] BUIJS, J. C. A. M.; VAN DONGEN, B. F.; VAN DER AALST, W. M. P. **Quality Dimensions in Process Discovery: the Importance of Fitness, Precision, Generalization and Simplicity**. *International Journal of Cooperative Information Systems*, 23(01), 2013.
- [7] BURATTIN, A. **{PLG2:} Multiperspective Process Randomization with Online and Offline Simulations**. In: *Proceedings of the {BPM} Demo Track 2016 Co-located with the 14th International Conference on Business Process Management {(BPM} 2016)*, Rio de Janeiro, Brazil, September 21, 2016., p. 1–6, 2016.
- [8] BURATTIN, A.; SPERDUTI, A. **Heuristics miner for time interval**. *Lecture Notes in Business Information Processing*, 207(April):85–95, 2010.
- [9] DE MEDEIROS, A. K. A. **Genetic Process Mining**. Phd, Eindhoven University of Technology, 2006.
- [10] DE MEDEIROS, A. K. A.; VAN DONGEN, B. F.; VAN DER AALST, W.; WEIJTERS, A. J. M. M. **Process mining for ubiquitous mobile systems: An overview and a concrete algorithm**. *Ubiquitous Mobile Information and Collaboration Systems (UMICS 2004)*, 3272(1):151–165, 2004.

- [11] DE MEDEIROS, A. K. A.; WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P. **Genetic process mining: An experimental evaluation.** *Data Mining and Knowledge Discovery*, 14(2):245–304, 2007.
- [12] DE SMEDT, J.; DE WEERDT, J.; VANTHIENEN, J. **Fusion Miner: Process discovery for mixed-paradigm models.** *Decision Support Systems*, 77:123–136, 2015.
- [13] DE WEERDT, J.; DE BACKER, M.; VANTHIENEN, J.; BAESENS, B. **A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs.** *Information Systems*, 37(7):654–676, 2012.
- [14] DI CICCIO, C.; MECELLA, M. **A two-step fast algorithm for the automated discovery of declarative workflows.** *Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, (i):135–142, 2013.
- [15] GÜNTHER, C. W.; VAN DER AALST, W. M. **Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics.** *Business Process Management - Lecture Notes in Computer Science*, 4714:328–343, 2007.
- [16] HALL, M.; FRANK, E.; GEOFFREY HOLMES, B. P.; REUTEMANN, P.; WITTEN, I. H. **The WEKA Data Mining Software: An Update.** *SIGKDD Explorations*, 11(1), 2009.
- [17] LEEMANS, S. J. J.; FAHLAND, D.; VAN DER AALST, W. M. P. **Discovering block-structured process models from event logs - A Constructive Approach.** In: *International Conference on Applications and Theory of Petri Nets and Concurrency*. Springer Berlin Heidelberg, 2013.
- [18] LI, J.; LIU, D.; YANG, B. **Process Mining: Extending α -Algorithm to Mine Duplicate Tasks in Process Logs.** *Advances in Web and Network Technologies, and Information Management*, p. 396–407, 2007.
- [19] MAGGI, F. M.; MOOIJ, A. J.; VAN DER AALST, W. M. P. **User-guided discovery of declarative process models.** *IEEE SSCI 2011: Symposium Series on Computational Intelligence - CIDM 2011: 2011 IEEE Symposium on Computational Intelligence and Data Mining*, p. 192–199, 2011.
- [20] MARUSTER, L.; WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P.; VAN DEN BOSCH, A. **A rule-based approach for process discovery: Dealing with noise and imbalance in process logs.** *Data Mining and Knowledge Discovery*, 13(1):67–87, 2006.

- [21] MUNOZ-GAMA, J.; CARMONA, J.; VAN DER AALST, W. M. P. **Conformance checking in the large: Partitioning and topology**. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8094 LNCS:130–145, 2013.
- [22] RIBEIRO, J.; CARMONA, J.; MISIR, M.; SEBAG, M. **A Recommender System for Process Discovery**. In: *International Conference on Business Process Management*. Springer International Publishing, 2014.
- [23] ROZINAT, A.; DE MEDEIROS, A. K. A.; GÜNTHER, C.; WEIJTERS, A.; VAN DER AALST, W. **Towards an evaluation framework for process mining algorithms**. *BPM Center Report BPM-07-06, BPMcenter.org*, 2007.
- [24] VAN DER AALST, W. M. P. **Process Mining: Discovery, Conformance and Enhancement of Business Processes**. Springer Berlin Heidelberg, 2011.
- [25] VAN DER AALST, W. M. P.; RUBIN, V.; VERBEEK, H. M. W.; VAN DONGEN, B. F.; KINDLER, E.; GÜNTHER, C. W. **Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting**. *Software and Systems Modeling*, 9(1):87–111, 2010.
- [26] VAN DER AALST, W. M. P. **Process Mining: Data Science in Action**. Springer, 2016.
- [27] VAN DER AALST, W. M. P.; PESIC, M.; SCHONENBERG, H. **Declarative Workflows - Balancing Between Flexibility and Support**. *Informatik - Forschung und Entwicklung*, p. 1–15, 2009.
- [28] VAN DER AALST, W. M.; TER HOFSTEDE, A.; KIEPUSZEWSKI, B.; BARROS, A. **Workflow Patterns**. *Distributed and Parallel Databases*, 14(1):5–51, 2003.
- [29] VAN DER AALST, W. M. P.; WEIJTERS, A. J. M. M.; MARUSTER, L. **Workflow Mining: Which Processes can be Rediscovered**. Technical report, Eindhoven University of Technology, Eindhoven, 2002.
- [30] VAN DER WERF, J. **Process discovery using integer linear programming**. In: *International Conference on Applications and Theory of Petri Nets*, p. 368–387. Springer, 2008.
- [31] VAN DONGEN, B. F. **Real-life event logs - Hospital log**, 2011. Eindhoven University of Technology, <https://doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffc54>.
- [32] VAN DONGEN, B. F. **BPI Challenge 2012**, 2012. Eindhoven University of Technology, <https://doi.org/10.4121/uuid:3926db30-f712-4394-aebe-75976070e91f>.

- [33] VANDEN BROUCKE, S. K. L. M. **Advances in Process Mining: Artificial Negative Events and Other Techniques**, 2014. Citeseer.
- [34] VANDEN BROUCKE, S. K. L. M.; DE WEERDT, J.; VANTHIENEN, J.; BAESENS, B. **A comprehensive benchmarking framework (CoBeFra) for conformance analysis between procedural process models and event logs in ProM**. *Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2013 - 2013 IEEE Symposium Series on Computational Intelligence, SSCI 2013*, p. 254–261, 2013.
- [35] VÁZQUEZ-BARREIROS, B.; MUCIENTES, M.; LAMA, M. **A genetic algorithm for process discovery guided by completeness, precision and simplicity**. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8659 LNCS:118–133, 2014.
- [36] WANG, J.; WONG, R.; DING, J.; J.A, W.; R.K.B, W.; J.A, D.; Q.A, G.; L.A, W.; WANG, J.; WONG, R.; DING, J. **Efficient selection of process mining algorithms**. *IEEE Transactions on Services Computing*, 6(4):484–496, 2013.
- [37] WEIJTERS, A. J. M. M.; RIBEIRO, J. **Flexible heuristics miner (FHM)**. In: *Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on*, número IEEE, 2011.
- [38] WEIJTERS, A. J. M. M.; VAN DER AALST, W. M. P.; DE MEDEIROS, A. K. A. **Process Mining with the HeuristicsMiner Algorithm**. *Cirp Annals-manufacturing Technology*, 166:1–34, 2006.
- [39] WEN, L.; VAN DER AALST, W. M. P.; WANG, J.; SUN, J. **Mining process models with non-free-choice constructs**. *Data Mining and Knowledge Discovery*, 15(2):145–180, 2007.
- [40] WEN, L.; WANG, J.; SUN, J.-G. **Mining Invisible Tasks from Event Logs**. *Advances in Data and Web Management*, p. 358–365, 2007.
- [41] WEN, L.; WANG, J.; VAN DER AALST, W. M. P.; HUANG, B.; SUN, J. **A novel approach for process mining based on event types**. *Journal of Intelligent Information Systems*, 32(2):163–190, 2008.
- [42] WESTERGAARD, M.; STAHL, C.; REIJERS, H. A. **UnconstrainedMiner: Efficient Discovery of Generalized Declarative Process Models**. *BPM Center Report, No. BPM-13-28*, 2013.

- [43] ZELST, S. J. V.; VAN DONGEN, B. F.; VAN DER AALST, W. M. P. **Filter Techniques for Region-Based Process Discovery**. Technical report, Technical Report 15–4, BPM Center. org, 2015.