

# ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ В ЗАДАЧАХ ИССЛЕДОВАНИЯ И МОДЕЛИРОВАНИЯ БИЗНЕС-ПРОЦЕССОВ

## DATA MINING IN BUSINESS PROCESS RESEARCH AND MODELING TASKS

**K. Dumbrays  
A. Dvoynova**

*Summary.* This thesis presents a survey of the tools of process mining to improve the effectiveness of the process approach to management in the organization. Information system collecting messages from the organization's information systems, converting them into a single business process event log, and then extracting the restored and real business process model from the event logs is being constructed. With the help of the process modeling based on various algorithms, it is possible to study several process parameters — repetition rate, execution time, cycles and various anomalies.

*Keywords:* process mining, event log, process detection, quality of the model.

**Думбрайс Крыстина Ольгертовна**

К.б.н., доцент, Международный университет  
«Дубна», г. Дубна  
2533208@mail.ru

**Двойнова Анастасия Дмитриевна**

Международный университет «Дубна», г. Дубна

*Аннотация.* В статье представлен обзор подходов интеллектуального анализа для повышения эффективности внедрения процессного подхода в организацию. Описана информационная система для сбора сообщений из информационных систем организации, преобразование их в единый журнал событий бизнес-процессов, а затем извлечение из журналов событий восстановленной и реальной модели бизнес-процесса. С помощью полученных моделей процессов можно изучить такие параметры процессов как частота повторений, время исполнения, циклы и аномалии в происходящих процессах.

*Ключевые слова:* интеллектуальный анализ процессов, журнал событий, выявление процессов, качество модели.

**П**роцессный подход к управлению организацией был разработан и применяется с целью создания горизонтальных связей в организациях. Подразделения и сотрудники, которые задействованы в одном процессе, могут самостоятельно координировать работу в рамках процесса и решать возникающие проблемы без участия вышестоящего руководства. Процессный подход к управлению позволяет более оперативно и эффективно решать возникающие вопросы и воздействовать на результат. В отличие от функционального подхода, управление процессами позволяет концентрироваться не на работе каждого из подразделений, а на результатах работы организации в целом. Основным элементом становится процесс. В соответствии с одним из принципов процессного подхода организация состоит не из подразделений, а из процессов.

Для выявления фактического бизнес-процесса, как и для его последующего совершенствования, часто используется графическое или текстовое описание на базе информации, полученной в ходе интервью с участниками процесса. Несмотря на всю простоту данного подхода, ключевым его недостатком является сложность извлечения знаний из участников процесса, что часто вызвано их нежеланием делиться информацией об особенностях процесса, и главное, о недостатках в своей работе.

В связи с этим, интеллектуальный анализ бизнес-процессов организаций становится весьма актуальным для повышения эффективности внедрения процессного подхода посредством создания информационной системы, основанной на технологии *process mining*.

*Process mining* представляет собой особую технику анализа бизнес-процессов, нацеленную на решение трёх задач: выявление или обнаружение процессов, проверка соответствия и расширение модели [1]. Начало анализа идёт не от моделирования к реализации, а процесс исследуется сначала с внутренней стороны. При этом речь идёт не только об автоматическом построении модели процесса в виде диаграммы, но и об определении его количественных характеристик. Свое начало интеллектуальный анализ берет с журнала событий, предоставляющего данные для анализа, на основе которых создаётся будущая модель бизнес-процесса на основе какого-либо алгоритма.

Обязательные поля журнала событий для *process mining* (рис. 1):

- ♦ *case ID* — *id* трассировки — набора событий, относящихся к отдельному конкретному экземпляру процесса;

id	activity	timestamp
3	register request	2010-12-30 14:32:00+01:00
3	examine casually	2010-12-30 15:06:00+01:00
3	check ticket	2010-12-30 16:34:00+01:00
3	decide	2011-01-06 09:18:00+01:00
3	reinitiate request	2011-01-06 12:18:00+01:00
3	examine thoroughly	2011-01-06 13:06:00+01:00
3	check ticket	2011-01-08 11:43:00+01:00
3	decide	2011-01-09 09:55:00+01:00
3	pay compensation	2011-01-15 10:45:00+01:00
2	register request	2010-12-30 11:32:00+01:00
2	check ticket	2010-12-30 12:12:00+01:00
2	examine casually	2010-12-30 14:16:00+01:00
2	decide	2011-01-05 11:22:00+01:00
2	pay compensation	2011-01-08 12:05:00+01:00
1	register request	2010-12-30 11:02:00+01:00
1	examine thoroughly	2010-12-31 10:06:00+01:00
1	check ticket	2011-01-05 15:12:00+01:00

Рис. 1. Фрагмент журнала событий

- ♦ *activity name* — название активности, то есть отдельного события;
- ♦ *timestamp* — время фиксации события в журнале событий. Может указываться также два значения: время начала события и время окончания события, благодаря чему удастся вычислить длительность события и задержки между событиями.

Дополнительные поля в журнале событий для осуществления интеллектуального анализа могут включать в себя различную справочную информацию — например, имя сотрудника, который отвечал за осуществление события, стоимость произошедшего события или наименование системы, откуда поступила информация о событии.

#### Основные методы интеллектуального анализа бизнес-процессов

В общем виде алгоритм проведения анализа по технологии *process mining* состоит из следующих этапов:

- ♦ *Discovery* (обнаружение).
- ♦ *Conformance checking* (проверка соответствия).
- ♦ *Enhancement* (улучшение).
- ♦ *Monitoring* (отслеживание).

Задача этапа проверки соответствия — определить, в какой мере фактический процесс совпадает с эталонным, выявить критические отклонения, мешающие за-

планированному течению процедур, провести своеобразное сравнение «ожидание/реальность».

Реализуется проверка соответствия реального процесса эталонному, регламентированному в той или иной компании. Первоначально происходит воссоздание реального процесса согласно следующим принципам [2]:

- ♦ выявление фактической, а не «экспертно-идеальной» последовательности действий;
- ♦ обнаружение повторяющихся и стандартных операций;
- ♦ выявление процента соответствия процессов эталонному пути;
- ♦ обнаружение «правильных путей» — последовательностей событий, чаще других приводящих к желаемому результату;
- ♦ поиск шаблонов поведения: циклы, «пинг-понг» и пр.

Основные задачи проверки соответствия:

- ♦ Увидеть «реальное» состояние бизнес-процесса в формате графической или аналитической модели;
- ♦ Сравнить «реальное» состояние и целевое состояние, зафиксированное в регламентах;
- ♦ Найти «узкие места» в бизнес-процессах;
- ♦ Выделить для последующего выборочного анализа «экстремальные» экземпляры процесса;

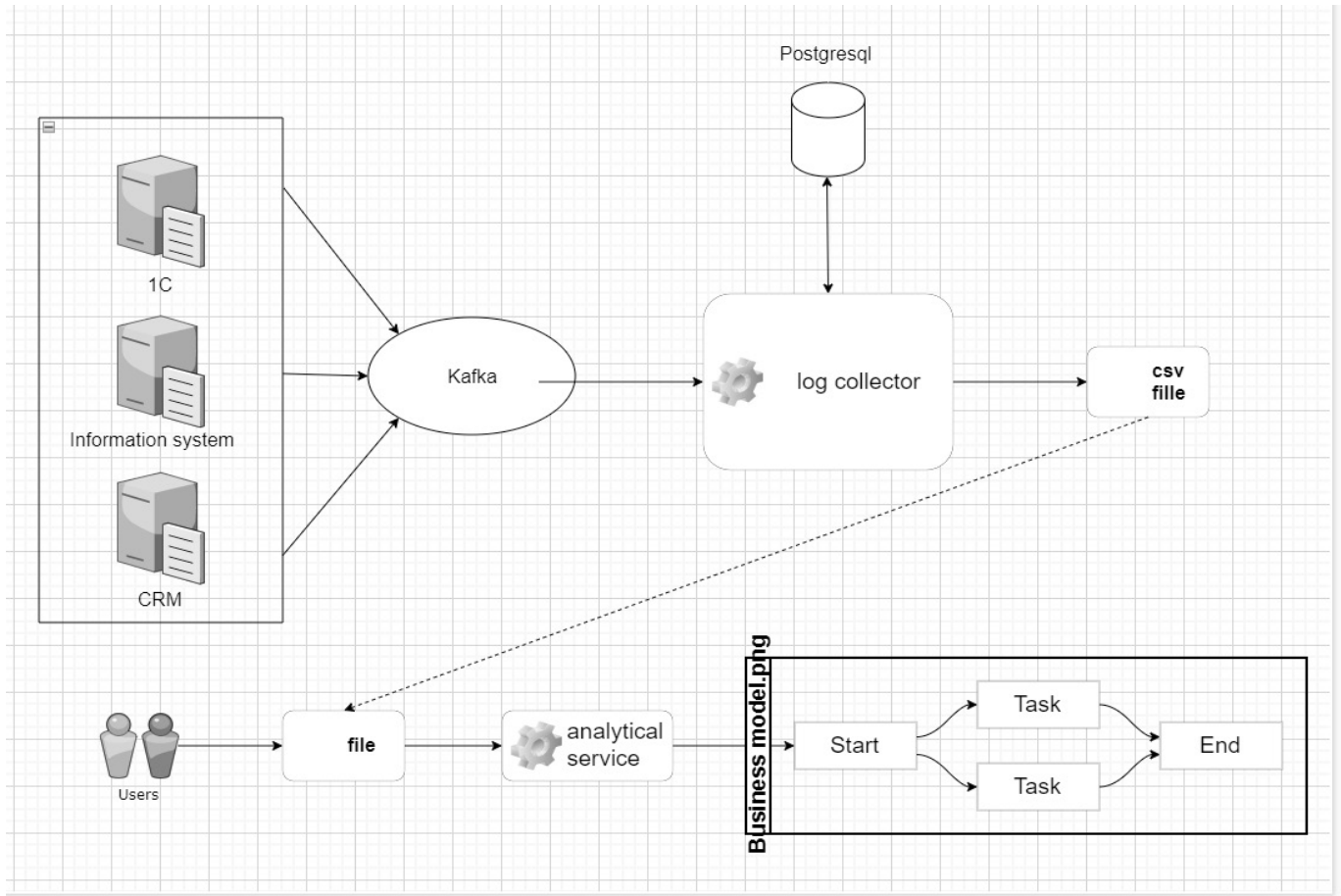


Рис. 2. Взаимодействие сервисов информационной системы

- ◆ Проанализировать процессные KPI;
- ◆ Проверить соответствие фактических параметров стандартам, заложенным в проекте (в первую очередь проверка длительности процессов).

Выявлять расхождения между смоделированным и наблюдаемым поведением помогают алгоритмы проверки соответствия. Они выдают показатели степени соответствия и диагностические сведения, объясняющие наблюдаемые различия. С их помощью можно детально анализировать прецеденты, не соответствующие построенной модели.

Соответствие модели определяется только для конкретного журнала событий. Для проведения проверки соответствия требуется, помимо лога событий, некоторая заранее определенная модель. Это модель может быть построена вручную или получена с помощью методов извлечения. Для расчета соответствия журнал событий последовательно исполняется моделью. Устанавливается начальная разметка модели (один маркер в начальной позиции). Алгоритм последовательно про-

ходит каждую трассу журнала. Для каждого очередного события возможны два варианта:

- ◆ переход с соответствующим именем активен и может работать
- ◆ нет активного перехода с соответствующим именем.

В первом случае переход срабатывает, разметка модели меняется, а алгоритм переходит к следующему событию журнала. Во втором случае в некоторых из позиций, необходимых для срабатывания соответствующего перехода модели, нет достаточного количества маркеров. В эту позицию добавляется искусственный маркер, который позволяет продолжить исполнение модели. Позиция помечается как содержащая соответствующее количество пропавших маркеров. Если при дальнейшем исполнении снова возникает подобная ситуация, количество пропавших маркеров в рассматриваемой позиции увеличивается. Возможна также ситуация, что при достижении финальной разметки в сети остаются маркеры, которые не были потреблены в ходе исполнения. Позиции, в которых эти маркеры расположены, помечаются как содержащие лишние маркеры.

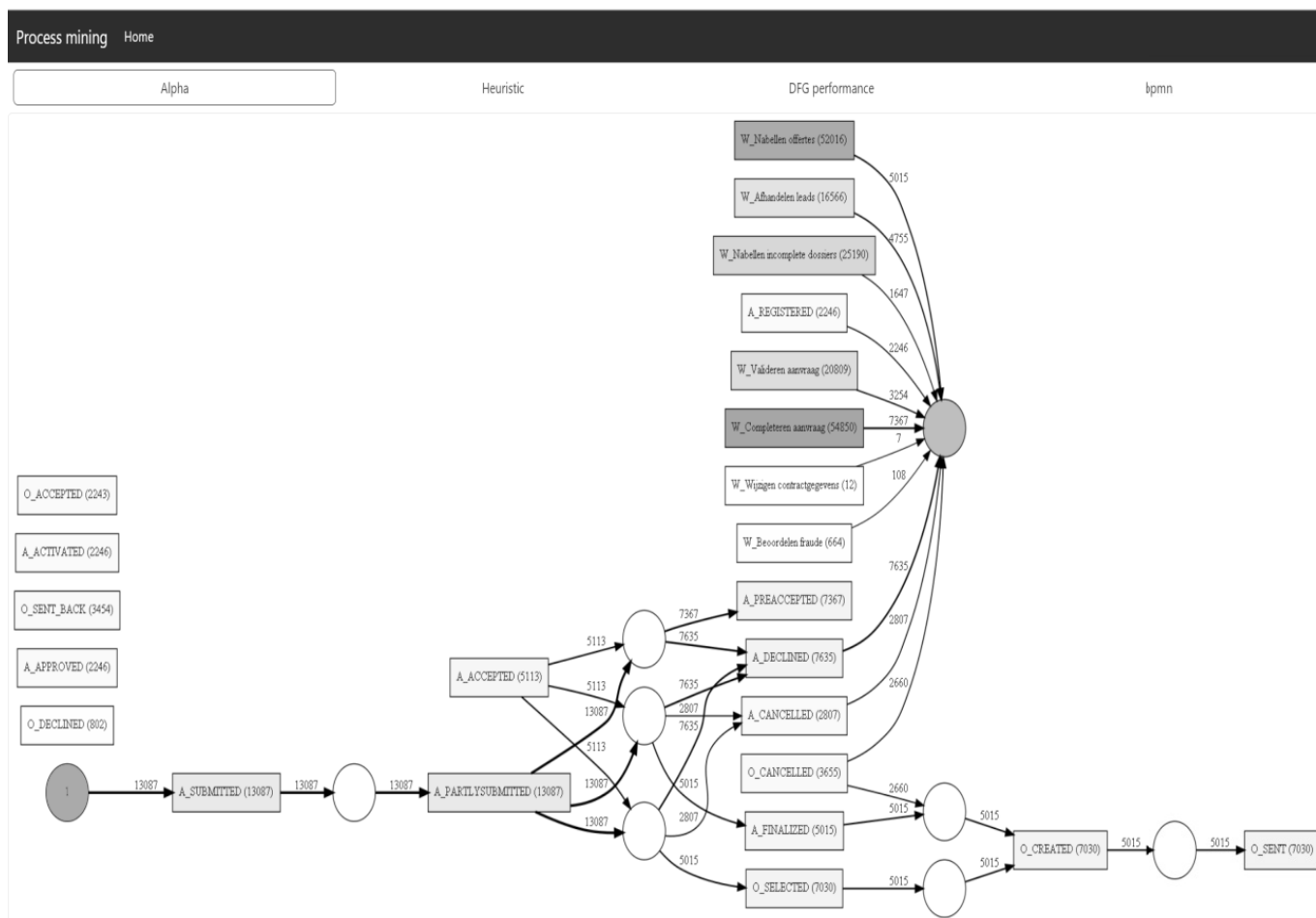


Рис. 3. Модель бизнес-процесса, построенная на основе альфа-алгоритма

В [3] соответствие модели  $N$  и журнала событий  $L$  определяется следующим образом:

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

где  $m_{\sigma}$  — количество пропавших маркеров,  $r_{\sigma}$  — количество лишних маркеров,  $c_{\sigma}$  — количество потребленных маркеров,  $p_{\sigma}$  — количество произведенных маркеров,  $\sigma$  — конкретная трасса. Существуют и другие, более совершенные способы оценки соответствия модели и журнала событий, например, в методике оценки соответствия на основе использования так называемых выравниваний [4].

Для создания информационной системы, состоящей из сервисов обработки и преобразования файла логов

в бизнес-модель, основанной на технологии *process mining*, были сформулированы следующие наборы требований:

**Бизнес-требования:**

- ◆ восстановление модели «as-is» бизнес-процесса без траты времени и ресурсов на интервьюирование сотрудников организации;
- ◆ оптимизация бизнес-процессов в организации по заданным критериям;

**Функциональные требования:**

- ◆ При взаимодействии с информационной системой пользователь имеет возможность загружать собственные файлы с данными о логах бизнес-процесса в различных форматах (*csv*, *xes*);
- ◆ Входными данными для подсистемы являются файлы с данными о бизнес-процессе, которые должны содержать обязательные столбцы, выходными данными — модели процессов, построенные с использованием различных алгоритмов.

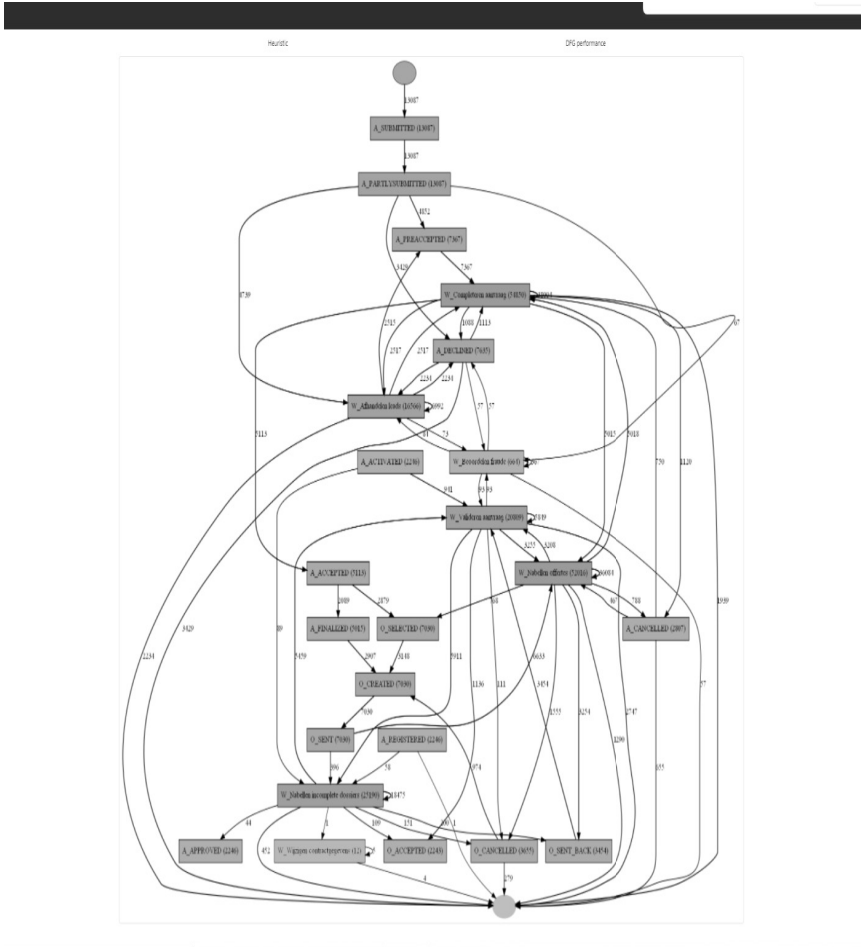


Рис. 4. Модель, построенная на эвристическом алгоритме

Наибольший интерес с практической точки зрения представляет аналитический сервис, обрабатывающий файлы двух форматов:

- ◆ CSV (*comma-separated values*)
- ◆ XES

XES — принятый повсеместно для технологии *process mining* формат логов, написанный на языке XML.

Модель, построенная на альфа-алгоритме, позволяет посмотреть, какие события в процессе происходят чаще всего (рис. 3). Это может быть полезно для понимания основной линии процесса, так и для нахождения каких-либо аномалий, бутылочных горлышек или ненужных циклов.

Модель, построенная на эвристическом алгоритме (рис. 4), так же визуально дает понять, какие события происходят часто, а какие редко (и нужны ли эти действия процесса). Кроме того, сразу можно заметить, что

семь событий процесса имеют цикличность, что может быть проблемой в достижении эффективности происходящего процесса.

Модель, построенная на графе прямого следования, кажется более обобщенной с большим количеством развилок, но дает пользователю информацию о времени выполнения действия процесса, что является очень важным показателем с точки зрения оптимизации процесса (рис. 5).

На модели сразу видно событие, которое выполняется месяц, возможно, этот шаг процесса необходимо оптимизировать и понять причины, приводящие к данному феномену.

Оценка качества полученных моделей

Под параметрами качества модели процесса понимаются соответствие (пригодность), простота, точность

Таблица 1. Результаты анализа качества модели

Показатель	Точность	Пригодность	Простота	Обобщенность
Результат	0.116146055848002	1.0	0.56097560975609	0.96142689987551

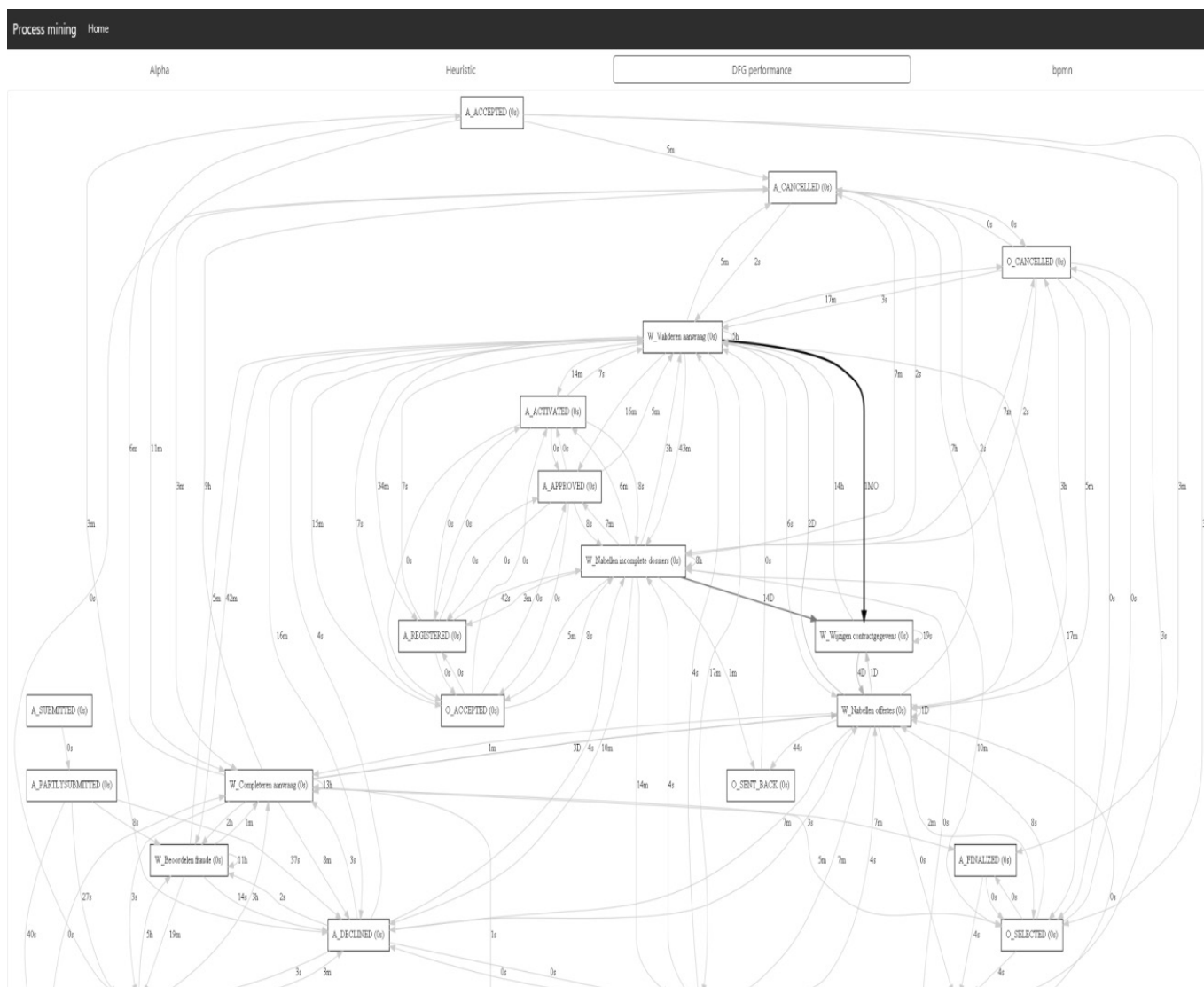


Рис. 5. Граф прямого следования с параметром времени выполнения

и обобщенность. Соответствие модели оценивается по тому, какова доля трасс событий воспроизводима моделью. Значение доли оценивается экспертом. Простота модели оценивается по тому, насколько она хорошо соответствует журналу событий и каким количеством узлов (позиций и переходов) и дуг, их связывающих, она представляется. Большое количество элементов модели говорит о ее запутанности и сложности

интерпретации. Под точностью модели понимается ее способность не допускать иного поведения, кроме изначально описанного. Обобщенность модели демонстрирует способность создавать различные варианты поведения, позволяя на этой основе прогнозировать процессы. Точность и обобщенность — два противоположных аспекта, требующие применения в комплексе и соблюдения баланса.

Для измерения точности полученной модели сети Петри [5] относительно журнала событий будет рассчитано с помощью первого из двух подходов:

- ◆ Соответствие требованиям
- ◆ Соответствие выравниванию и т.д. (с использованием выравниваний)

Идея, лежащая в основе этих двух подходов, одна и та же: разные префиксы журнала воспроизводятся (если это возможно) в модели. При достижении отметки набор переходов, которые разрешены в модели процесса, сравнивается с набором действий, которые следуют за префиксом. Чем больше наборов отличаются друг от друга, тем ниже значение точности. Чем больше наборов похоже, тем выше значение точности. Это справедливо только в том случае, если работает повтор префикса в модели процесса: если повтор не дает результата, префикс не учитывается при вычислении точности. Следовательно, точность, рассчитанная поверх непригодных процессов, на самом деле не имеет смысла. Основное различие между этими подходами заключается в методе воспроизведения. Воспроизведение на основе токенов происходит быстрее, но основано на эвристике (следовательно, результат воспроизведения может быть неточным). Выравнивания точны,

но могут быть медленными, если пространство состояний велико.

В таблице 1 приведены показатели, полученные в результате анализа качества модели:

Показатели качества модели позволяют сделать вывод о том, что модель пригодна для использования, довольно проста, но, в то же время, является сильно обобщенной. Можно попробовать очистить журнал от шума или лишних и неполных трас с целью улучшения некоторых показателей.

Оптимизация всех описанных параметров качества одновременно, как правило, невозможна. Улучшение одних параметров качества может ухудшать другие, например, модели с высоким соответствием могут иметь низкие показатели по точности и обобщенности. Однако возможности различных алгоритмов обеспечивать параметры качества необходимо учитывать при приложении их к реальным процессам. Вероятность достижения оптимальных параметров качества зависит от конструкций реальных процессов, а также степени зашумленности и неполноты журнала.

#### ЛИТЕРАТУРА

1. Process Mining (PM) — процессная аналитика [Электронный ресурс]. — Режим доступа: <https://Processmi-com.turbopages.org/Processmi.com/s/terms/Process-Mining/>. — Дата доступа: 09.06.2022.
2. Processmining: шаг за шагом — Электрон. текст. — 2022. — Режим доступа: <https://loginom.ru/blog/process-mining-step-step> (дата обращения: 20.05.2022)
3. Rozinat A. Process Mining: Conformance and Extension. PhD Thesis, Eindhoven University of Technology. 2020.
4. Мицюк А.А. Синтез моделей процессов по журналам событий с шумом / А.А. Мицюк, И.С. Шугуров // Моделирование и анализ информационных систем. — 2014.
5. Абросимова М.А. Алгоритмы обнаружения процессов в process mining / М.А. Абросимова // Современные инструментальные системы, информационные технологии и инновации: Сборник научных трудов XVII Международной научно-практической конференции, Курск, 17–18 марта 2022 года / Редколлегия: Разумов М.С. (отв. ред.). — Курск: Юго-Западный государственный университет, 2022. — С. 31–39. — EDN DCPZNV.

© Думбрайс Крыстына Ольгертовна ( 2533208@mail.ru ), Двойнова Анастасия Дмитриевна.

Журнал «Современная наука: актуальные проблемы теории и практики»