

DOI: 10.17323/2587-814X.2024.3.41.55

# Скрытая марковская модель: метод построения модели бизнес-процесса\*

**А.Ю. Варнухов**

E-mail: varnuhov\_ayu@usue.ru

Уральский государственный экономический университет, Екатеринбург, Россия

## Аннотация

Все больше компаний под воздействием стремительного развития технологий (концепция Индустрия 4.0/5.0) охватывают процессы цифровой трансформации. Внедрение информационных систем обеспечивает возможность накопления большого объема данных о деятельности компании. Исследование такой информации расширяет возможности применения data-driven подхода к управлению бизнес-процессами (business process management, BPM). Обработка и изучение данных из журналов событий с помощью методов интеллектуального анализа процессов позволяет строить цифровые модели бизнес-процессов, которые оказываются полезным источником сведений при проведении работ по анализу, моделированию и реинжинирингу в рамках процессного подхода. В настоящей работе разрабатывается метод построения модели бизнес-процесса на основе скрытой марковской модели с учетом ограничений, налагаемых предметной областью. Применение скрытой марковской модели позволяет использовать аппарат теории вероятностей и математической статистики для анализа бизнес-процессов, а также решать задачи классификации и кластеризации. В статье описываются возможности data-driven подхода к управлению бизнес-процессами и демонстрируются примеры практического применения метода для решения бизнес-задач: построение графа зависимостей, который может быть использован для выявления расхождений между фактическим и ожидаемым исполнением, а также способ предсказания исхода бизнес-процесса на основе последовательности наблюдаемых событий.

**Ключевые слова:** бизнес-процессы, скрытые марковские модели, интеллектуальный анализ процессов, бизнес-анализ, предсказание, классификация, data-driven подход, информационные системы, журналы событий

**Цитирование:** Варнухов А.Ю. Скрытая марковская модель: метод построения модели бизнес-процесса // Бизнес-информатика. 2024. Т. 18. № 3. С. 41–55. DOI: 10.17323/2587-814X.2024.3.41.55

\* Статья опубликована при поддержке Программы НИУ ВШЭ «Университетское партнерство»

## Введение

Развитие возможностей современных информационных технологий стимулирует предприятия различных сфер переводить свои бизнес-процессы из «аналоговой» формы в цифровую. Существует и постоянно совершенствуется множество методологий и методик, которые позволяют выполнять моделирование, проводить реинжиниринг, контролировать и отслеживать бизнес-процессы [1]. Довольно часто моделирование выполняется «вручную» с привлечением соответствующих бизнес-аналитиков и «внутренних» экспертов, которые обладают специальными знаниями о моделируемых явлениях. При этом на практике процесс моделирования и реинжиниринга бизнес-процессов оказывается нетривиальной задачей даже для опытных специалистов [2]. Так, например, имеют место искажения, обусловленные субъективными факторами, собственным положением в структуре организации и другими типичными проблемами, характерными для такого подхода к моделированию: идеализацией, выбором неверного уровня абстракции или неспособностью адекватно воспроизводить наблюдаемое взаимодействие [3]. В результате полученная модель может отражать только часть происходящей «реальности», оказывается недостаточно функциональной и, в конечном счете, будет обладать весьма ограниченной ценностью.

Внедрение автоматизированных информационных систем различного класса и функциональности (ERP, CRM, ECM и прочие) приводит к сопут-

ствующему накоплению в цифровых хранилищах большого объема полезных сведений о деятельности предприятия [4]. Обработка и последующий анализ данных, накопленных в информационных системах предприятия, обеспечивает возможность применения data-driven подхода. В настоящее время ведутся исследования в области интеллектуального анализа процессов [5, 6], создания цифровых двойников [7], предиктивной и прескриптивной аналитики [8, 9], роботизированной автоматизации процессов [10], а также проводятся работы по практическому внедрению результатов этих исследований в различных отраслях [11, 12].

## 1. Применение data-driven подхода к управлению бизнес-процессами

Процессный подход позволяет представить организацию в виде совокупности взаимосвязанных бизнес-процессов, каждый из которых рассматривается как ценный актив, обеспечивающий поставку продуктов и услуг компании конечным потребителям. Методология **управления бизнес-процессами** (business process management, BPM) определяет жизненный цикл управления, который, как правило, состоит из следующих основных этапов: анализа, моделирования, исполнения, мониторинга, оптимизации и реинжиниринга. Для формулирования возможностей и контекста применения data-driven подхода в рамках BPM можно построить обобщенный жизненный цикл управления, показанный на *рисунке 1*.



Рис. 1. Обобщенный жизненный цикл BPM.

Как видно из *рисунка 1*, модели процессов являются источником информации при проведении анализа и оптимизации, оказывают поддержку на этапе внедрения информационных систем и способствуют реализации функций управления и контроля. Накопленные в информационных системах данные могут быть использованы для создания цифровых моделей, которые позволят лучше понимать реальные бизнес-процессы организации. Применение data-driven подхода открывает возможность установления тесной взаимосвязи между реальными процессами и их представлением в виде моделей. Среди множества вариантов применения можно выделить несколько основных: обнаружение моделей, проверка соответствия, оценка показателей и усовершенствование процесса. Для обнаружения моделей исследуются данные из журналов событий информационных систем и с помощью специальных методов интеллектуального анализа производится построение моделей без привлечения какой-либо априорной информации. Полученные таким образом модели «AS-IS» могут быть использованы в дальнейших работах при проведении анализа, моделирования и реинжиниринга. Стоит отметить, что этап построения является крайне важным, поскольку обеспечивает своего рода «входную» точку для всех остальных задач, применяющих обнаруженные цифровые модели. Для проверки соответствия используются ранее построенная эталонная модель бизнес-процесса и данные из журнала событий, которые сравниваются между собой с целью определения степени соответствия. Такая проверка оказывается полезной для контроля соблюдения

накладываемых правил и ограничений, выявления расхождений между фактическим и ожидаемым исполнением, поиска причин отклонения и так далее. Если принять во внимание наличие в данных временной компоненты, то с помощью моделей можно проводить измерение показателей производительности бизнес-процесса, обнаруживать «узкие места», оценивать уровень сервисного обслуживания и так далее. Так, например, вариантный анализ позволит обнаружить различия в потоке управления и показателях производительности между различными подразделениями организации. Кроме исследования потока управления, можно расширить модель путем включения организационной составляющей, что позволит учитывать информацию об участниках процесса и их взаимосвязях. Таким образом применение data-driven подхода позволяет повысить качество и эффективность управления бизнес-процессами.

## 2. Постановка задачи для построения модели бизнес-процесса

Если бизнес-процесс можно представить в виде модели, то конкретный отдельный случай, реализуемый в рамках этой модели, можно описать как ее экземпляр. Экземпляр бизнес-процесса должен характеризоваться некоторым набором из последовательных или параллельных действий (деятельностей) с возможностью определения порядка их следования. Разные экземпляры должны быть различимы между собой хотя бы порядком следования событий. В *таблице 1* представлен пример фрагмента журнала событий, полученного из информационной системы.

Таблица 1.

Фрагмент журнала событий

ID Экземпляра	ID События	Временная метка		Событие	Сотрудник	...
1001	24837	24.08.2023	13:20	Поступление запроса	Иванов А.	...
1001	25123	25.08.2023	11:05	Проверка наличия	Петрова И.	...
1001	26001	26.08.2023	09:15	Отправка счета	Иванов А.	...
1001	26560	27.08.2023	16:07	Отгрузка товара	Сидоров В.	...
1002	24842	24.08.2023	14:27	Поступление запроса	Иванов А.	...
1002	24859	24.08.2023	16:20	Проверка наличия	Петрова И.	...
1002	24892	24.08.2023	17:40	Отказ в поставке	Соболев Б.	...
...	...	...	...	...	...	...

Подразумевается, что все данные в журнале относятся к одному анализируемому бизнес-процессу. Каждая строка в таблице содержит следующие обязательные атрибуты: «ID экземпляра», «Событие» и «Временная метка». Множество строк с одинаковым значением атрибута «ID экземпляра» представляют события, которые связаны с одним экземпляром бизнес-процесса. Атрибут «Событие» содержит название события, которое можно связать с некоторым действием (деятельностью). Атрибут «Временная метка» используется для хронологического упорядочивания событий в рамках одного экземпляра. Журнал бизнес-процесса может содержать и другие дополнительные атрибуты («Сотрудник», «Стоимость», «Клиент», «Офис» и так далее), которые могут быть полезны для мониторинга бизнес-процесса с применением машинного обучения [13]. Для краткости записи будем использовать мультимножество, которое будет состоять из хронологически упорядоченных и сгруппированных последовательностей событий согласно журналу бизнес-процесса. Например, для данных, приведенных в таблице 1, можно записать мультимножество:

$$L = \{ \langle a, b, c, d \rangle^n, \langle a, b, e \rangle^m, \dots \}, \quad (1)$$

где  $L$  – мультимножество, в котором каждый элемент содержит упорядоченную последовательность событий:  $a$  – событие «Поступление запроса»,  $b$  – событие «Проверка наличия»,  $c$  – событие «Отправка счета»,  $d$  – событие «Отгрузка товара»,  $e$  – событие «Отказ в поставке»;  $n$  и  $m$  – количество раз, которые встречалась данная упорядоченная последовательность в журнале.

Таким образом, требуется разработать метод построения модели бизнес-процесса на основе поступающих входных данных в виде мультимножества  $L$ .

### 3. Анализ методов построения модели бизнес-процесса

**«Альфа-искатель».** Является довольно простым и одним из первых методов, который позволяет воссоздать модель бизнес-процесса из имеющегося набора последовательных событий в виде Workflow-net (частный случай сети Петри) [14]. Для этого алгоритм сканирует журнал в поисках определенного набора паттернов: последовательности, XOR разделения и AND разделения. На основе

этого записывается матрица «отпечатков», позволяющая распознать имеющиеся отношения между событиями. По этой матрице, с учетом правил вывода, строится конечная модель. Среди ограничений «альфа-искателя» можно отметить: сложности при обработке зашумленных данных, невозможность распознать циклы длиной 1 или 2 шага, проблемы с локальными зависимостями.

**«Эвристический искатель».** В отличие от «альфа-искателя» применяет идею подсчета частот появления событий и воспроизводит модель процесса в виде причинно-следственной сети (Causal net) [15, 16]. В начале рассчитываются метрики, которые оценивают количество связей прямого следования между каждой парой событий и измеряют степень их зависимости. При помощи паттернов (последовательности, XOR, AND и цикла), опирающихся на рассчитанные метрики, строится граф зависимости (Dependency graph). Поиск слияний и разделений в графе зависимостей может выполняться скользящим окном по журналу событий с заданным размером или на основе решения оптимизационной задачи, в которой целевой функцией является степень соответствия модели наблюдаемому журналу. Полученная модель процесса в виде Causal net может быть преобразована к другим требуемым нотациям (BPMN, UML, EPC, WF-net и так далее). Этот метод менее восприимчив к зашумленности в данных и устраняет многие недостатки «альфа-искателя», однако имеет проблемы с обработкой не-локальных зависимостей и обнаружением дублирующих событий, а также требует ручной настройки уровней порога отсека.

**«Искатель на основе областей».** Базируется на применении теории регионов и строится на допущении, что модели состояний могут быть преобразованы в сети Петри [17]. Существует несколько подходов реализации этого метода. Первый подход заключается в определении области как множества таких состояний, что действия в модели состояний и переходов согласуется с данной областью. В этом случае все события можно разделить на «входящие», «исходящие» и «внутренние» по отношению к данной области. После разделения областей по этим правилам каждая область может быть сопоставлена с определенной позицией в сети Петри. Второй подход применяет специально заданную языковую модель вместо системы состояний и переходов [18]. Основная идея подхода заключается в том, что удаление  $P_i$  позиции не приводит к

удалению любого поведенческого паттерна, однако добавление новой позиции может привести к устранению некоторых возможных вариантов поведения. К достоинствам можно отнести возможность обрабатывать более комплексные структуры потока управления. Слабой стороной этого метода является неспособность обнаруживать некоторые типы конструкций процесса, проблемы с точностью и обобщающей способностью, а также сложность его практической реализации.

**«Индуктивный искатель».** Состоит из трех рекурсивно выполняемых шагов: построение ориентированного графа, поиск среза и разделение записей журнала [19]. В качестве исходных данных метод использует предварительно подготовленный журнал событий. На первом шаге метод преобразует данные в ориентированный граф, в котором каждый узел соответствует одному событию, а дуги формируют переходы между событиями. После этого выполняется попытка обнаружить места возможных срезов. Если места таких срезов обнаруживаются, то алгоритм образует оператор среза и сегменты разделения. На основе обнаруженных сегментов производится декомпозиция журнала на более мелкие составляющие. Затем происходит рекурсивная обработка каждого полученного таким образом фрагмента до обнаружения базового случая: фрагмент содержит только одно событие. Если в процессе рекурсивного спуска встречается фрагмент, который не сводится к базовому случаю и при этом не имеет допустимых мест для среза, то применяется процесс «проваливания». Базовая реализация метода имела сложности с обнаружением циклов фиксированной длины, обработкой редких событий и ограничений, связанных с рекурсивной природой построения. Однако дальнейшее развитие метода позволило преодолеть первичные недостатки, обеспечило возможность масштабирования и применения распределенных вычислений [20].

Представленные выше методы позволяют построить модель бизнес-процесса различными способами, однако интерес представляет исследование возможности построения модели, которая основана на оценке вероятностей. В качестве основы можно было бы рассмотреть цепь Маркова, но, учитывая характер и природу исходных данных, более приемлемым будет предположение, что зафиксированные в журнале события являются только внешним проявлением некоторого скрытого от наблюдателя процесса. Для моделирования такого

предположения можно рассмотреть скрытую марковскую модель (далее – СММ) первого порядка. Известно, что такие модели со скрытыми состояниями эффективно применяются для задач обработки текста на естественных языках [21], идентификации жестов [22], распознавания речи [23], биоинформатике [24] и других сферах. Исходя из сведений, которые представлены в проанализированных источниках, можно ожидать, что применение СММ для анализа бизнес-процесса сделает возможным не только построение модели бизнес-процесса, но и позволит решать задачу классификации и выполнять кластеризацию данных.

#### 4. Предлагаемый метод построения модели

СММ обладает множеством скрытых состояний  $S = \{s_1, s_2, s_3, \dots, s_N\}$ . Каждое скрытое состояние может быть связано с некоторыми другими скрытыми состояниями. Схематичное представление модели показано на рисунке 2.

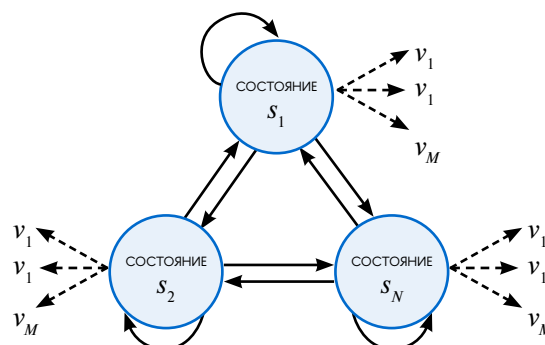


Рис. 2. Схематичное представление СММ.

В данной работе рассматривается полносвязная структурная организация СММ, при которой каждое скрытое состояние  $s_k$  связано со всеми отличными от него скрытыми состояниями, а также с самим собой. Помимо скрытых состояний определен конечный алфавит множества наблюдаемых событий  $V = \{v_1, v_2, v_3, \dots, v_M\}$  и каждое скрытое состояние воспроизводит события из данного множества  $V$ . В любой отдельный момент времени  $t$  модель находится в одном из скрытых состояний.

$$\forall t: q_t \in S, 1 \leq t \leq T. \quad (2)$$

СММ совершает переходы между скрытыми состояниями. Так, в момент времени  $t$ , находясь в

скрытом состоянии  $q_t$ , модель перейдет в другое состояние с определенной вероятностью и в момент времени  $t + 1$  будет находиться в скрытом состоянии  $q_{t+1} \in S$ . В данной работе рассматриваются только дискретные моменты времени, при этом текущее состояние и цепочка выполненных переходов между ними являются невидимыми для наблюдателя. Находясь в некотором скрытом состоянии  $q_t$ , модель воспроизводит событие  $o_t \in V$ , которое видно внешнему наблюдателю. Череда переходов между состояниями и воспроизведенными ими событиями в результате формирует последовательность наблюдений  $O = \{o_1, o_2, o_3, \dots, o_T\}$ . Схематичное представление работы СММ показано на *рисунке 3*.

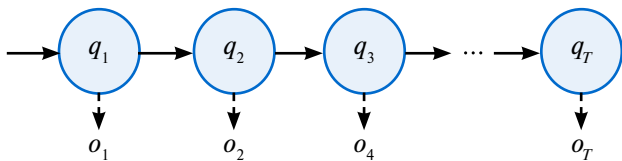


Рис. 3. Схематичное представление работы СММ.

Поскольку в данной работе рассматривается СММ первого порядка, то согласно марковскому свойству будем полагать, что вероятность перехода из одного состояния в другое определяется только предыдущим состоянием модели:

$$\begin{aligned} P(q_t = s_j | q_1 = s_n, q_2 = s_m, \dots, q_{t-1} = s_j) = \\ = P(q_t = s_j | q_{t-1} = s_j). \end{aligned} \quad (3)$$

Второе допущение будет заключаться в том, что вероятность произвести наблюдаемое событие  $o_t$  зависит только от состояния, в котором находится модель в дискретный момент времени  $t$  и не зависит от других состояний и наблюдаемых событий:

$$\begin{aligned} P(o_t = v_k | q_1 = s_n, \dots, q_t = s_i, o_1 = v_m, \dots, o_{t-1} = v_j) = \\ = P(o_t = v_k | q_t = s_i). \end{aligned} \quad (4)$$

Определим начальное распределение по скрытым состояниям модели, которое задает вероятность того, что модель на первом шаге будет находиться в некотором состоянии:

$$\pi = \{\pi_i\}_{i=1}^N, \quad \pi_i = P(q_1 = s_i), \quad \sum_{i=1}^N \pi_i = 1. \quad (5)$$

Определим распределение вероятностей перехода между скрытыми состояниями как матрицу  $A = (a_{ij})$ , где

$$a_{ij} = P(q_t = s_j | q_{t-1} = s_i), \quad 1 \leq i, j \leq N, \quad \sum_{j=1}^N a_{ij} = 1. \quad (6)$$

Распределение вероятностей появления событий при нахождении модели в некотором скрытом состоянии определим как матрицу  $B = (b_{ik})$ , где

$$\begin{aligned} b_{ik} = P(o_t = v_k | q_t = s_i), \quad 1 \leq i \leq N, \quad 1 \leq k \leq M, \\ \sum_{k=1}^M b_{ik} = 1. \end{aligned} \quad (7)$$

Исходя из вышеописанного скрытую марковскую модель  $\theta$  определим как

$$\theta = (S, V, A, B, \pi). \quad (8)$$

Предположим, что представлен исходный журнал и выполнена предварительная подготовка данных. Пусть задано мультимножество

$$\begin{aligned} L = \{ \langle a, e \rangle^5, \langle a, b, c, e \rangle^5, \langle a, c, b, e \rangle^5, \\ \langle a, d, e \rangle^{10}, \langle a, d, d, e \rangle^5, \langle a, d, d, d, e \rangle^1, \\ \langle a, b, c, d, e \rangle^3, \langle a, b, d, c, e \rangle^3, \langle a, c, b, d, e \rangle^3, \\ \langle a, c, d, b, e \rangle^3, \langle a, d, c, b, e \rangle^3, \langle a, d, b, c, e \rangle^3, \\ \langle e, b, c, m \rangle^4, \langle e, c, b, m \rangle^3 \}. \end{aligned} \quad (9)$$

Мультимножество  $L$  содержит повторяющиеся несколько раз элементы, которые представляют отдельные экземпляры бизнес-процесса, выполненного в разное время. Можно заметить, что некоторые элементы мультимножества (например,  $\langle a, b, c, e \rangle$  и  $\langle a, c, b, e \rangle$ ) содержат почти одинаковые последовательности событий за тем исключением, что порядок, в котором следуют события «b» и «c», переставлен местами. Видимая перестановка в мультимножестве может встречаться из-за того, что зарегистрированные события попадают в исходный журнал упорядоченными по временной метке, однако в реальности представляют собой параллельно исполняемые подпроцессы бизнес-процесса. Пример такой ситуации показан на *рисунке 4*.

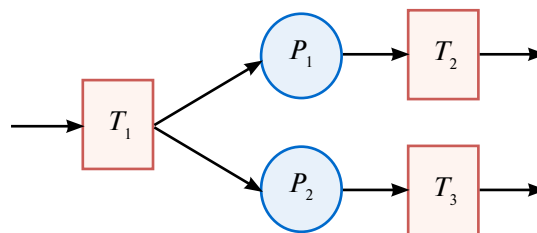


Рис. 4. Параллельно исполняемые подпроцессы.

В связи с тем, что такие события всегда встречаются попарно и следуют друг за другом (хоть и в разном порядке), а также учитывая характер моделируемой предметной области, будем полагать, что подобные переставленные местами события образуют логически единую операцию бизнес-процесса. Таким образом, сходные попарные перестановки событий, которые по своему поведению образуют логическую операцию «AND» бизнес-процесса, будем относить к одному и тому же скрытому состоянию модели. Обычно бизнес-процесс имеет одно фиксированное стартовое событие, в котором начинается его исполнение. Финальных событий у бизнес-процесса может быть несколько в силу необходимости представления различных результатов его завершения. Каждое такое финальное событие является логически завершающим и поэтому не должно быть разделимо на несколько параллельно исполняемых подпроцессов. Если параллельно исполняемые подпроцессы не находятся в стартовом или финальном событиях, то следует предположить, что существует некое событие, после которого происходит разветвление исполнения, равно как и событие, соединяющее параллельное исполнение. Таким образом, логическая операция «AND» бизнес-процесса должна находиться между некоторым начальным и конечным событием в наблюдаемой последовательности событий. Для определения групп событий, которые формируют множество логических операций «AND» бизнес-процесса выберем уникальные элементы по правилу (10) из мультимножества  $L$ :

$$FSET = \left\{ \begin{array}{l} (\sigma(i) = S_1, \sigma(i+1) = S_2, \sigma(i+2) = \\ = S_3, \sigma(i+3) = S_4): \\ \exists (\delta(l) = S_1, \delta(l+1) = S_3, \delta(l+2) = \\ = S_2, \delta(l+3) = S_4) \end{array} \right\}$$

$$\forall \sigma, \delta \in L : |\sigma| \geq 4, |\delta| \geq 4,$$

$$1 \leq i \leq |\sigma|, 1 \leq l \leq |\delta|. \quad (10)$$

Членами множества  $FSET$  являются упорядоченные последовательности событий, каждая из которых удовлетворяет следующим условиям:

- ◆ получена из входящих в исходное мультимножества  $L$  элементов;
- ◆ последовательность содержит минимум четыре идущих по порядку события;
- ◆ существует другая четверка с таким же началь-

ным событием  $S_1$  и конечным событием  $S_4$  в которой события  $S_2$  и  $S_3$  переставлены местами.

Для рассматриваемого мультимножества из (9) будет сформировано множество  $FSET$ , состоящее из следующих элементов:

$$FSET(L) = \{(a, b, c, e), (a, c, b, e), (a, b, c, d), \\ (b, c, d, e), (a, b, d, c), (b, d, c, e), (a, c, b, d), (c, b, d, e), \\ (a, c, d, b), (c, d, b, e), (a, d, c, b), (d, c, b, e), \\ (a, d, b, c), (d, b, c, e), (e, b, c, m), (e, c, b, m)\}. \quad (11)$$

Каждый элемент множества  $FSET$  представляет собой минимально допустимую часть возможной перестановки. Так, например, для образования логической операции «AND», составленной из двух параллельно выполняемых подпроцессов « $b$ » и « $c$ », которая начинается после события « $a$ » и заканчивается событием « $e$ », необходимо, чтобы множество  $FSET$  содержало обе части такой перестановки  $(a, b, c, e)$  и  $(a, c, b, e)$ . Элементы множества  $(a, b, d, c)$  и  $(a, d, b, c)$  также образуют минимальную логическую операцию «AND». Однако из (9) можно заметить, что элементы « $b$ » и « $d$ » являются частью большей логической операции «AND», которая включает также и событие « $c$ ». Таким образом, необходимо определить процедуру наращивания более длинных логических операций «AND», состоящих из базовых минимальных частей. Для этой цели зададим множества стартовых, перестановочных и конечных событий:

$$FS = \{\sigma(1) : \sigma \in FSET\}. \quad (12)$$

$$PS = \{\sigma(2) : \sigma \in FSET\}. \quad (13)$$

$$ES = \{\sigma(4) : \sigma \in FSET\} \setminus PS. \quad (14)$$

Для наращивания максимально возможной перестановки поочередно возьмем каждое стартовое событие из множества  $FS$  и перебирая множество элементов  $FSET$  будем добавлять каждые встреченные перестановочные символы во второй и третьей позиции, пока не дойдем до элемента, который содержит конечное событие из  $ES$  в последний позиции. Процедура наращивания показана в листинге 1.

В результате процедуры наращивания будет сформировано мультимножество, которое содержит элементы с выделенными начальным и конечным состояниями, а также множеством перестановочных событий между ними. Для примера из (9) сформируется мультимножество

**Input:** *FS*: Start events  
**Input:** *ES*: End events  
**Input:** *FSET*: Set of data  
**Output:** *FPERM*: Multiset of augmented permutations

1.  $FPERM \leftarrow \emptyset$
2. **for** each start event *SS* **in** *FS* **do**
3.  $ACCPERM \leftarrow \emptyset$
4. **for** each element  $\sigma$  **in** *FSET* **do**
5.     **if**  $\sigma(1) = SS$  **or**  $|ACCPERM| > 0$  **then**
6.          $ACCPERM \leftarrow ACCPERM \cup \sigma(2) \cup \sigma(3)$
7.     **if**  $|ACCPERM| > 0$  **and**  $\sigma(4) \in ES$  **then**
8.          $FPERM \leftarrow FPERM \uplus (SS, ACCPERM, \sigma(4))$
9.      $ACCPERM \leftarrow \emptyset$

Листинг 1. Процедура наращивания.

$$FPRM(L) = \{ \langle e, \{b, c\}, m \rangle^2, \langle a, \{b, c\}, e \rangle^2, \langle a, \{b, c, d\}, e \rangle^6, \langle b, \{c, d\}, e \rangle^2, \langle c, \{b, d\}, e \rangle^2, \langle d, \{b, c\}, e \rangle^2 \}. \quad (15)$$

Кратность элементов в мультимножестве *FPERM* отражает частоту, с которой данная перестановка была встречена в исходных данных. Допустим, что в исходных данных присутствовала группа событий «*b*», «*c*», «*d*», которая образует логическую операцию «AND» бизнес-процесса. Тогда кратность такой группы с одинаковыми начальными и конечными событиями должна быть равной шести. Предположим, что в этой группе отсутствует один элемент, например (*a*, *c*, *b*, *d*, *e*). В этом случае группа событий должна распасться на две логические операции «AND» формирующие параллельные подпроцессы: первая группа событий начинается с события «*b*» и включает «*c*» и «*d*», вторая начинается с события «*d*» и включает «*c*» и «*b*». Возможна также ситуация с потерей некоторых данных в процессе выгрузки или их предварительной обработки. Для учета этих ситуаций введем метрику контроля:

$$\text{MinLimit}(\sigma) = |\sigma(2)| - \varepsilon, \sigma \in FPERM, \varepsilon \geq 0. \quad (16)$$

Метрика (16) позволяет рассчитать необходимое количество одинаковых элементов с учетом возможности корректировки на недостающие или пропавшие данные. Сформируем множество уникальных групп событий, образующих логические операции «AND», путем включения в него только элементов кратность которых не менее заданного предела:

$$LPERM = \left\{ \sigma : \sum_{s \in FPERM} [s = \sigma] \geq \text{MinLimit}(\sigma) \right\}, \quad \sigma \in FPERM. \quad (17)$$

В получившемся множестве кроме наращенных элементов содержатся также части большей перестановки. Для исключения подобных лишних элементов выполним проверку:

$$ANDGROUP = \left\{ \sigma : \nexists \delta \mid \delta(2) = \{\sigma(1)\} \cup \sigma(2) \wedge \delta(3) = \sigma(3) \right\}, \sigma, \delta \in LPERM. \quad (18)$$

Таким образом множество *ANDGROUP* будет содержать только необходимые группы событий, которые представляют параллельные исполняемые подпроцессы бизнес-процесса. Для примера из (9) получим следующие логические операции «AND»:

$$ANDGROUP(L) = \{ \langle e, \{b, c\}, m \rangle, \langle a, \{b, c\}, e \rangle, \langle a, \{b, c, d\}, e \rangle \}. \quad (19)$$

Из (8) следует, что для определения СММ требуется задать множества скрытых состояний и событий, а также определить матрицы переходов, эмиссии и вектор вероятностей, характеризующий выбор начального состояния. Для произвольного бизнес-процесса никакие из этих параметров модели заранее не известны, поскольку имеется только наблюдаемая последовательность событий, которая получена из исходных данных. Таким образом, требуется сформулировать метод, позволяющий найти параметры, если известна только последовательность наблюдаемых событий бизнес-процесса. Данная проблема была сформулиро-



вана в работе Л. Рабинера и является одной из трех базовых проблем при работе СММ и одновременно самой сложной из них [25]. Сложность обуславливается отсутствием известных аналитических способов решения задачи, позволяющих определить параметры модели для любой конечной последовательности  $O$ . Существуют несколько подходов ее решения путем сведения проблемы к задаче оптимизации для поиска таких параметров модели  $\theta$ , позволяющих максимизировать вероятность  $P(O|\theta)$ . Одним из таких подходов является алгоритм Баума–Велча, который является разновидностью EM (expectation-maximization) алгоритма поиска оценок наибольшего правдоподобия. В общем виде этот алгоритм состоит из двух шагов (E-шага и M-шага), позволяющих итеративно выполнять пересчет параметров  $\theta$  и последовательно приближаться к локально максимальной оценке при определенном  $O$ .

Однако классическая реализация алгоритма Баума–Велча не учитывает особенности предметной области и специфику функционирования бизнес-процессов. Поэтому в данной работе предлагается доработанная для применения к рассматриваемой задаче модификация алгоритма.

Определим множество наблюдаемых событий  $V$  модели как равное множеству уникальных событий бизнес-процесса из мультимножества  $L$ :

$$V = \bigcup_{\sigma \in L} \{ \sigma(i), 1 \leq i \leq |\sigma| \}. \quad (20)$$

Поскольку каждый элемент в (18) представляет собой операцию логического «AND» и любая уникальная группа перестановочных событий должна быть отнесена к одному скрытому состоянию модели, зададим множество скрытых состояний:

$$SPU = \{ \sigma(2) : \sigma \in ANDGROUP \}. \quad (21)$$

$$SOU = V \setminus \bigcup \delta, \forall \delta \in SPU. \quad (22)$$

$$S = \{ i : 1 \leq i \leq |SOU| + |SPU| \}. \quad (23)$$

Итеративная природа реализации классического алгоритма Баума–Велча допускает задание матриц  $A$  и  $B$  произвольными величинами перед началом его работы, поскольку в процессе обновления параметров будет достигнута сходимость к оптимальным значениям. Однако известно, что различные структуры организации СММ (эргодическая, левосторонняя, параллельная левосторон-

няя и так далее) могут оказывать влияние на характер ее поведения и получаемые результаты. Пусть каждое событие из множества  $V$  пронумеровано по порядку натуральным числом  $v_k$  от 1 до  $|V|$ . Определим матрицы  $A$  и  $B$  следующим образом:

$$a_{ij} = \frac{1}{|S|}, 1 \leq i, j \leq |S|. \quad (24)$$

$$b_{ik} = \begin{cases} 1 : \forall i \in \{ 1 \leq i \leq |SOU| \} \wedge \forall k = \\ = v_k \in SOU \mid \forall i \neq j \ b_j \neq b_i \\ \frac{1}{|\sigma|} : \forall i \in \{ |SOU| < i \leq |S| \} \wedge \forall k = \\ = v_k \in \sigma \mid \forall i \neq j \ b_j \neq b_i, \forall \sigma \in SPU \\ 0 : \text{иначе.} \end{cases} \quad (25)$$

Для задания начального распределения будем полагать, что стартовое событие бизнес-процесса единственно и для него определено  $v_k = 1$ :

$$\pi_i = \begin{cases} 1 : i = v_k \\ 0 : i \neq v_k \end{cases}. \quad (26)$$

Если в мультимножестве  $L$  существуют элементы, которые содержат различные стартовые события, то всегда возможно добавить в начало всех элементов мультимножества новое суррогатное событие, чтобы перейти к единственности стартового события.

Для снижения количества операций и упрощения расчетов применяется метод прямого и обратного прохода, который основан на принципах динамического программирования. При этом формируется матрица промежуточных значений, позволяющая на каждом шаге оценивать вероятность путем суммирования вычислений, произведенных на предыдущих шагах, посредством вспомогательных функций:

$$\alpha_t(i) = P(o_1, o_2, o_3, \dots, o_t, q_t = s_i | \theta). \quad (27)$$

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, o_{t+3}, \dots, o_T | q_t = s_i, \theta). \quad (28)$$

В классической реализации (27) и (28) не учитываются особенности модели, связанные со спецификой рассматриваемой задачи. Необходимо принять в расчет ограничения, которые налагаются на переходы между скрытыми состояниями, соотношенными с операциями логического «AND». Для этого определим вспомогательные функции следующим образом:

$$\varphi(i) = \{v_k | b_{ik} > 0, 1 \leq k \leq |V|\}. \quad (29)$$

$$HFW_t(i) = \begin{cases} |\varphi(i)| \times \prod_{j=0}^{|\varphi(i)|-1} [o_{t+j} \in \varphi(i)] : T > t + |\varphi(i)| \\ 0 : T \leq t + |\varphi(i)|. \end{cases} \quad (30)$$

Функция (30) задает оценку, которая будет использоваться для выбора наиболее подходящего скрытого состояния при вычислении  $\alpha$  и  $\beta$ . Для такого состояния значение функции (30) будем максимальным:

$$HFW_t^{\max} = \operatorname{argmax}_{1 \leq i \leq N} HFW_t(i). \quad (31)$$

Тогда будем иметь:

$$\alpha_1(i) = \pi_i b_i(o_1). \quad (32)$$

$$\alpha_{t+1}(j) = \begin{cases} b_j(o_{t+1}) \sum_{i=1}^N \alpha_t(i) a_{ij} : (j = HFW_{t+1}^{\max} \wedge \\ \wedge o_{t+1} \in PS \wedge o_t \notin PS) \vee (o_{t+1} \notin PS) \\ b_j(o_{t+1}) \alpha_t(j) a_{jj} : o_{t+1} \in PS. \end{cases} \quad (33)$$

Такое определение (33) позволяет ограничить пространство переходов между скрытыми состояниями модели, которые содержат группу событий, образующих операцию логического «AND». Похожим образом определим:

$$HBW_t(i) = \begin{cases} |\varphi(i)| \times \prod_{j=0}^{|\varphi(i)|-1} [o_{t-j} \in \varphi(i)] : t - |\varphi(i)| \geq 0 \\ 0 : t - |\varphi(i)| < 0 \end{cases} \quad (34)$$

$$HBW_t^{\max} = \operatorname{argmax}_{1 \leq i \leq N} HBW_t(i). \quad (35)$$

$$\beta_T(i) = 1. \quad (36)$$

$$\beta_t(i) = \begin{cases} \sum_{j=1}^N \beta_{t+1}(j) a_{ij} b_j(o_{t+1}) : (i = HBW_t^{\max} \wedge \\ \wedge o_t \in PS \wedge o_{t+1} \notin PS) \vee (o_t \notin PS) \\ \beta_{t+1}(i) b_i(o_{t+1}) a_{ii} : o_t \in PS \\ 0 : o_t \in PS \wedge o_{t+1} \notin PS \wedge i \neq HBW_t^{\max}. \end{cases} \quad (37)$$

Поскольку для бизнес-процесса финальное событие означает его полное завершение и невозможность перехода в любые другие состояния, а также для соответствия (7), будем полагать, что такие скрытые со-

стояния должны переходить сами в себя образуя петлю. Определим  $\xi$  и  $\gamma$  следующим образом:

$$\xi_t(i, j) = \begin{cases} \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} : 1 \leq t \leq T-1 \\ 1 : t = T \wedge i = j \wedge \alpha_{T-1}(i) > 0. \end{cases} \quad (38)$$

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{j=1}^N \alpha_t(j) \beta_t(j)}. \quad (39)$$

Учитывая, что обновление коэффициентов должно выполняться по всем элементам мультимножества  $L$  будем использовать:

$$a_{ij}^* = \frac{\sum_{\sigma \in L} \sum_{t=1}^T \xi_t(i, j)}{\sum_{\sigma \in L} \sum_{t=1}^T \gamma_t(i)}. \quad (40)$$

$$b_{ik}^* = \frac{\sum_{\sigma \in L} \sum_{t=1}^T [o_t = v_k] \gamma_t(i)}{\sum_{\sigma \in L} \sum_{t=1}^T \gamma_t(i)}. \quad (41)$$

Итеративное выполнение E и M шагов алгоритма выполняется до его сходимости, либо до достижения заданного предельного количества повторений.

## 5. Варианты применения метода для решения бизнес-задач

### 5.1. Предсказание исхода бизнес-процесса и поиск отклонений

Допустим, имеется множество экземпляров бизнес-процесса, которые разделены по некоторой характеристике на несколько непересекающихся групп  $G_1, G_2, G_3, \dots, G_N$ . Например, в рамках бизнес-процесса «продажа товара», можно провести разбиение его экземпляров по исходу сделки. В этом случае могут быть сформированы следующие группы: «отказались от покупки», «отложили покупку» и «успешно совершили покупку». Каждая такая группа соответствует своему мультимножеству  $L_1, L_2, L_3, \dots, L_N$ . Построим по вышеописанному предлагаемому методу  $N$  скрытых марковских моделей, используя эти мультимножества в качестве обучающей выборки. В результате каждому  $L_n$  будет соответствовать  $\theta_n$ . Воспользуемся алгоритмом прямого-обратного прохода и определим (42):

$$\alpha_t(j) = \begin{cases} \pi_j b_j(o_1) : t=1 \\ b_j(o_t) \sum_{i=1}^N \alpha_{t-1}(i) a_{ij} : 1 < t \leq T. \end{cases} \quad (42)$$

Для нового экземпляра бизнес-процесса  $O_x$  при помощи построенных СММ можно предсказать его принадлежность к одной из ранее сформированных групп. Группа  $G_n$ , для которой наиболее высока оценка  $P(O_x | \theta_n)$ , будет целевой:

$$F(O_x) = \operatorname{argmax}_{\theta} P(O_x | \theta) = \operatorname{argmax}_{\theta} \sum_{i=1}^N \alpha_T(i). \quad (43)$$

В результате экземпляр  $O_x$  с наибольшей вероятностью будет иметь исход, который соответствует группе  $G_n$ . Такое предсказание может быть получено и для незавершенных экземпляров бизнес-процесса, то есть для тех случаев, когда имеется только часть последовательности  $O_x$ . Обладая возможностью получения такой оценки, можно решать различные практические задачи. Например, для бизнес-процесса продажи товара можно провести анализ сделок, которые находятся на некотором промежуточном этапе, для прогнозирования возможного исхода. Если будет определена высокая вероятность нежелательного исхода, то для таких сделок можно выработать корректирующие меры воздействия, направленные на исправление траектории движения. Кроме этого, имея на входе «эталонную» модель бизнес-процесса и данные из соответствующего журнала событий можно, получив оценку принадлежности каждой последовательности, выявлять девиантные экземпляры с целью дальнейшего анализа причин и принятия управленческих решений.

### 5.2. Представление бизнес-процесса в виде графа зависимостей

Предположим, что для некоторого процесса, собран и обработан журнал событий, на основе которого сформировано мультимножество  $L$ :

$$L = \{ \langle a, z \rangle^5, \langle a, e, b, c, z \rangle^5, \langle a, e, c, b, z \rangle^3, \langle a, b, c, d, z \rangle^5, \langle a, b, d, c, z \rangle^5, \langle a, c, b, d, z \rangle^5, \langle a, c, d, b, z \rangle^5, \langle a, d, c, b, z \rangle^5, \langle a, d, b, c, z \rangle^5, \langle a, l, z \rangle^{10}, \langle a, l, l, z \rangle^5, \langle a, l, l, l, z \rangle^1 \}. \quad (44)$$

Если построить СММ для мультимножества (44) по вышеописанному предлагаемому методу, то получим следующие матрицы  $A$  и  $B$ :

$$A = \begin{pmatrix} 0 & 0 & 0,51 & 0,14 & 0,27 & 0,08 \\ 0 & 0,5 & 0 & 0 & 0 & 0,5 \\ 0 & 0 & 0,67 & 0 & 0 & 0,33 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0,3 & 0,7 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0,5 & 0,5 & 0 & 0 & 0 & 0 \\ 0 & 0,33 & 0,33 & 0,33 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Полагаем, что скрытые состояния, для которых  $a_{ii} > 0$  и которые содержат несколько событий в матрице  $B$ , соотносятся с операцией логического «AND», а имеющие только одно событие образуют цикл. Построенный граф зависимостей для мультимножества (44) показан на рисунке 5.

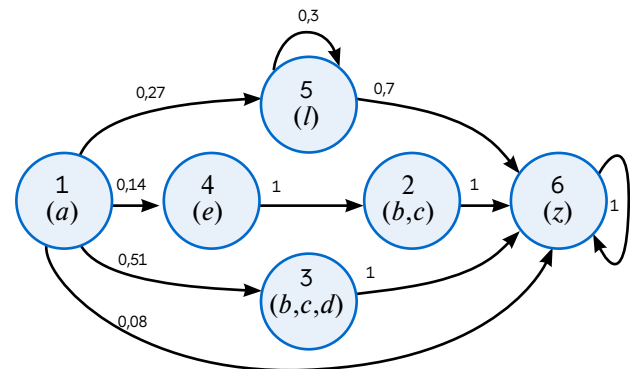


Рис. 5. Граф зависимостей бизнес-процесса.

При необходимости данный граф зависимостей может быть преобразован в другие представления: BPMN, сети Петри, Casual Net и так далее. Полученный граф зависимостей может быть использован для исследования фактического исполнения бизнес-процесса, проведения сравнительного анализа вариантов реализации между различными структурными подразделениями, поиска отклонений и выявления их причин. Если дополнить модель данными из журнала событий о времени выполнения базовых операций, то можно рассчитать различные показатели производительности

(время обработки и простоя, продолжительность и эффективное время одного цикла и так далее). Кроме того, в журналах событий может содержаться информация об участниках, понесенных затратах и используемых ресурсах, которая позволит расширить модель для анализа других аспектов бизнес-процесса.

### Заключение

Data-driven подход не является альтернативой традиционному моделированию с использованием аналитиков и специалистов в предметной области. Однако применение данного подхода позволяет улучшить качество проводимого анализа, моделирования, проектирования и реинжиниринга бизнес-процессов за счет исследования реальных данных, которые накоплены в информационных системах предприятия. Обнаружение неочевидных связей, а также возможность беспристрастного анализа, вне зависимости от субъективной точки зрения участников процесса, способствует минимизации вероятности появления искажений и ошибочных выводов. Построенная модель может применяться для контроля над исполнением конкретных экземпляров бизнес-процесса, выявления отклонений или нетипичного

поведения, а также обеспечит поддержку внедрения ключевых показателей эффективности (KPI), как на уровне отдельных исполнителей, так и для целых подразделений.

В отличие от других описанных в работе алгоритмов, предлагаемый метод основывается на скрытой марковской модели, которая позволяет использовать аппарат теории вероятностей и математической статистики. В частности, продемонстрирован способ получения оценки будущего исхода бизнес-процесса, который делает возможным реализацию упреждающего управленческого воздействия с целью корректировки ожидаемого результата. Помимо этого, при помощи СММ, можно выполнять кластеризацию экземпляров бизнес-процесса и решать задачу классификации.

К выявленным недостаткам относятся: отсутствие гарантированного появления всех событий, образующих логическую операцию «AND» (при использовании модели в качестве генератора), а также узкий горизонт учета зависимостей (из-за допущения первого порядка).

В качестве направления развития метода целесообразно рассмотреть многоуровневую иерархическую организацию модели, внедрение ансамблевых методов машинного обучения и применение СММ более высокого порядка. ■

### Литература

1. Lizano-Mora H., Palos-Sanchez P.R., Aguayo-Camacho M. The evolution of business process management: A bibliometric analysis // IEEE Access. 2021. Vol. 9. P. 51088–51105. <https://doi.org/10.1109/ACCESS.2021.3066340>
2. Fetais A., Abdella G.M., Al-Khalifa K.N., Hamouda A.M. Business process re-engineering: A literature review-based analysis of implementation measures // Information. 2022. Vol. 13. No. 4. Article 185. <https://doi.org/10.3390/info13040185>
3. Rosemann M. Potential pitfalls of process modeling: part A // Business Process Management Journal. 2006. Vol. 12. No. 2. P. 249–254. <https://doi.org/10.1108/14637150610657567>
4. Nambiar A., Mundra D. An overview of data warehouse and data lake in modern enterprise data management // Big Data and Cognitive Computing. 2022. Vol. 6. No. 4. Article 132. <https://doi.org/10.3390/bdcc6040132>
5. Pegoraro M., van der Aalst W.M.P. Mining uncertain event data in process mining // 2019 International Conference on Process Mining (ICPM). 2019. P. 89–96. <https://doi.org/10.1109/ICPM.2019.00023>
6. Andrews R., van Dun C.G.J., Wynn M.T., Kratsch W., Röglinger M.K.E., ter Hofstede A.H.M. Quality-informed semi-automated event log generation for process mining // Decision Support Systems. 2020. Vol. 132. Article 113265.
7. Park G., van der Aalst W.M.P. Realizing a digital twin of an organization using action-oriented process mining // 3rd International Conference on Process Mining (ICPM). 2021. P. 104–111. <https://doi.org/10.1109/ICPM53251.2021.9576846>
8. Kratsch W., Manderscheid J., Röglinger M. et al. Machine learning in business process monitoring: a comparison of deep learning and classical approaches used for outcome prediction // Business & Information Systems Engineering. 2021. Vol. 63. P. 261–276. <https://doi.org/10.1007/s12599-020-00645-0>
9. Teinimaa I., Dumas M., Rosa M.L., Maggi F.M. Outcome-oriented predictive process monitoring: Review and benchmark // ACM Transactions on Knowledge Discovery from Data (TKDD). 2019. Vol. 13. No. 2. P. 1–57. <https://doi.org/10.1145/3301300>

10. Leno V., Polyvyanyy A., Dumas M., La Rosa M., Maggi F.M. Robotic process mining: vision and challenges // *Business & Information Systems Engineering*. 2021. Vol. 63. P. 301–314. <https://doi.org/10.1007/s12599-020-00641-4>
11. Munoz-Gama J., Martin N., Fernandez-Llatas C. et al. Process mining for healthcare: Characteristics and challenges // *Journal of Biomedical Informatics*. 2022. Vol. 127. Article 103994. <https://doi.org/10.1016/j.jbi.2022.103994>
12. Grisold T., Mendling J., Otto M., vom Brocke J. Adoption, use and management of process mining in practice // *Business Process Management Journal*. 2021. Vol. 27. No. 2. P. 369–387. <https://doi.org/10.1108/BPMJ-03-2020-0112>
13. Mehdiyev N., Fettke P. Explainable artificial intelligence for process mining: A general overview and application of a novel local explanation approach for predictive process monitoring // *Interpretable artificial intelligence: A perspective of granular computing*. Springer, 2021. P. 1–28. [https://doi.org/10.1007/978-3-030-64949-4\\_1](https://doi.org/10.1007/978-3-030-64949-4_1)
14. van der Aalst W.M.P., Weijters T., Maruster L. Workflow mining: Discovering process models from event logs // *IEEE transactions on knowledge and data engineering*. 2004. Vol. 16. No. 9. P. 1128–1142. <https://doi.org/10.1109/TKDE.2004.47>
15. Mannhardt F., de Leoni M., Reijers H.A. Heuristic mining revamped: an interactive, data-aware, and conformance-aware miner // *15th International Conference on Business Process Management (BPM 2017)*. 2017. P. 1–5.
16. van der Aalst W.M.P., Adriansyah A., van Dongen B. Causal nets: A modeling language tailored towards process discovery // *International conference on concurrency theory*. 2011. P. 28–42. [https://doi.org/10.1007/978-3-642-23217-6\\_3](https://doi.org/10.1007/978-3-642-23217-6_3)
17. van Dongen B.F., Busi N., Pinna G.M., van der Aalst W.M.P. An iterative algorithm for applying the theory of regions in process mining // *Proceedings of the Workshop on Formal Approaches to Business Processes and Web Services (FABPWS'07)*. 2007. P. 36–55.
18. Bergenthum R., Desel J., Lorenz R., Mauser S. Process mining based on regions of languages // *Business Process Management: 5th International Conference (BPM 2007)*, Brisbane, Australia, September 24–28, 2007. P. 375–383. [https://doi.org/10.1007/978-3-540-75183-0\\_27](https://doi.org/10.1007/978-3-540-75183-0_27)
19. Leemans S.J.J., Fahland D., van der Aalst W.M.P. Discovering block-structured process models from event logs: A constructive approach // *Application and Theory of Petri Nets and Concurrency: 34th International Conference (PETRI NETS 2013)*, Milan, Italy, June 24–28, 2013. P. 311–329. [https://doi.org/10.1007/978-3-642-38697-8\\_17](https://doi.org/10.1007/978-3-642-38697-8_17)
20. Leemans S.J.J., Fahland D., van der Aalst W.M.P. Scalable process discovery with guarantees // *Enterprise, Business-Process and Information Systems Modeling (BPMDS EMMASD 2015)*. Lecture Notes in Business Information Processing. Vol. 214. Springer, Cham, 2015. P. 85–101. [https://doi.org/10.1007/978-3-319-19237-6\\_6](https://doi.org/10.1007/978-3-319-19237-6_6)
21. Pande S.D., Kanna R.K., Qureshi I. Natural language processing based on name entity with n-gram classifier machine learning process through ge-based hidden Markov model // *Machine Learning Applications in Engineering Education and Management*. 2022. Vol. 2. No. 1. P. 30–39.
22. Sagayam K.M., Hemanth D.J. A probabilistic model for state sequence analysis in hidden Markov model for hand gesture recognition // *Computational Intelligence*. 2019. Vol. 35. No. 1. P. 59–81. <https://doi.org/10.1111/coin.12188>
23. Srivastava R.K., Pandey D. Speech recognition using HMM and Soft Computing // *Materials Today: Proceedings*. 2022. Vol. 51. P. 1878–1883. <https://doi.org/10.1016/j.matpr.2021.10.097>
24. Du J., Wang C., Wang L. et al. Automatic block-wise genotype-phenotype association detection based on hidden Markov model // *BMC bioinformatics*. 2023. Vol. 24. Article 138. <https://doi.org/10.1186/s12859-023-05265-5>
25. Rabiner L.R. A tutorial on hidden Markov models and selected applications in speech recognition // *Proceedings of the IEEE*. 1989. Vol. 77. No. 2. P. 257–286. <https://doi.org/10.1109/5.18626>

### Об авторе

#### Варнухов Артем Юрьевич

ассистент, кафедра бизнес-информатики, Уральский государственный экономический университет, Россия, 620144, г. Екатеринбург, ул. 8 Марта, д. 62;

E-mail: varnuhov\_ayu@usue.ru

# Hidden Markov model: Method for building a business process model

**Artem Yu. Varnukhov**

E-mail: varnuhov\_ayu@usue.ru

Ural State University of Economics, Ekaterinburg, Russia

## Abstract

More and more companies are influenced by the rapid development of technology (Industry 4.0/5.0 concept), are embracing digital transformation processes. The introduction of information systems makes it possible to accumulate a large amount of data about the company's activities. Study of such information expands the opportunities for applying a data-driven approach to business process management (BPM). Processing and studying data from event logs using process mining methods make it possible to build digital models of business processes which turn out to be a useful source of information when carrying out analysis, modeling and reengineering within the framework of the process approach. In this paper, we develop a method for building a business process model based on a hidden Markov model, taking into account the restrictions imposed by the subject area. The use of a hidden Markov model allows us to use the apparatus of probability theory and mathematical statistics to analyze business processes, as well as to solve classification and clustering problems. This article describes the capabilities of a data-driven approach to business process management and demonstrates examples of the practical application of the method to solve business challenges: drawing a dependency graph that can be used to identify discrepancies between actual and expected execution, as well as a method for predicting the outcome of a business process based on the sequence of observed events.

**Keywords:** business processes, hidden Markov models, process mining, business analysis, prediction, classification, data-driven approach, information systems, event logs

**Citation:** Varnukhov A.Yu. (2024) Hidden Markov model: Method for building a business process model. *Business Informatics*, vol. 18, no. 3, pp. 41–55. DOI: 10.17323/2587-814X.2024.3.41.55

## References

1. Lizano-Mora H., Palos-Sánchez P.R., Aguayo-Camacho M. (2021) The evolution of business process management: A bibliometric analysis. *IEEE Access*, vol. 9, pp. 51088–51105. <https://doi.org/10.1109/ACCESS.2021.3066340>
2. Fetais A., Abdella G.M., Al-Khalifa K.N., Hamouda A.M. (2022) Business process re-engineering: A literature review-based analysis of implementation measures. *Information*, vol.13, no. 4, 185. <https://doi.org/10.3390/info13040185>
3. Rosemann M. (2006) Potential pitfalls of process modeling: part A. *Business Process Management Journal*, vol. 12, no. 2, pp. 249–254. <https://doi.org/10.1108/14637150610657567>
4. Nambiar A., Mundra D. (2022) An overview of data warehouse and data lake in modern enterprise data management. *Big Data and Cognitive Computing*, vol. 6, no. 4, 132. <https://doi.org/10.3390/bdcc6040132>
5. Pegoraro M., van der Aalst W.M.P. (2019). Mining uncertain event data in process mining. *2019 International Conference on Process Mining (ICPM)*, pp. 89–96. <https://doi.org/10.1109/ICPM.2019.00023>
6. Andrews R., van Dun C.G.J., Wynn M.T., Kratsch W., Röglinger M.K.E., ter Hofstede A.H.M. (2020) Quality-informed semi-automated event log generation for process mining. *Decision Support Systems*, vol. 132, 113265. <https://doi.org/10.1016/j.dss.2020.113265>

7. Park G., van der Aalst W.M.P. (2021) Realizing a digital twin of an organization using action-oriented process mining. *3rd International Conference on Process Mining (ICPM)*, pp. 104–111. <https://doi.org/10.1109/ICPM53251.2021.9576846>
8. Kratsch W., Manderscheid J., Röglinger M. et al. (2021) Machine learning in business process monitoring: A comparison of deep learning and classical approaches used for outcome prediction. *Business & Information Systems Engineering*, vol. 63, pp. 261–276. <https://doi.org/10.1007/s12599-020-00645-0>
9. Teinemia I., Dumas M., Rosa M.L., Maggi F.M. (2019) Outcome-oriented predictive process monitoring: Review and benchmark. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 13, no. 2, pp. 1–57. <https://doi.org/10.1145/3301300>
10. Leno V., Polyvyanyy A., Dumas M., La Rosa M., Maggi F.M. (2021) Robotic process mining: vision and challenges. *Business & Information Systems Engineering*, vol. 63, pp. 301–314. <https://doi.org/10.1007/s12599-020-00641-4>
11. Munoz-Gama J., Martin N., Fernandez-Llatas C. et al. (2022) Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, vol. 127, 103994. <https://doi.org/10.1016/j.jbi.2022.103994>
12. Grisold T., Mendling J., Otto M., vom Brocke J. (2021) Adoption, use and management of process mining in practice. *Business Process Management Journal*, vol. 27, no. 2, pp. 369–387. <https://doi.org/10.1108/BPMJ-03-2020-0112>
13. Mehdiyev N., Fettke P. (2021) Explainable artificial intelligence for process mining: A general overview and application of a novel local explanation approach for predictive process monitoring. *Interpretable Artificial Intelligence: A Perspective of Granular Computing*, Springer, pp. 1–28. [https://doi.org/10.1007/978-3-030-64949-4\\_1](https://doi.org/10.1007/978-3-030-64949-4_1)
14. van der Aalst W.M.P., Weijters T., Maruster L. (2004) Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, no. 9, pp. 1128–1142. <https://doi.org/10.1109/TKDE.2004.47>
15. Mannhardt F., de Leoni M., Reijers H.A. (2017) Heuristic mining revamped: an interactive, data-aware, and conformance-aware miner. *15th International Conference on Business Process Management (BPM 2017)*, pp. 1–5.
16. van der Aalst W.M.P., Adriansyah A., van Dongen B. (2011) Causal Nets: A modeling language tailored towards Process Discovery. *International Conference on Concurrency Theory*, pp. 28–42. [https://doi.org/10.1007/978-3-642-23217-6\\_3](https://doi.org/10.1007/978-3-642-23217-6_3)
17. van Dongen B.F., Busi N., Pinna G.M., van der Aalst W.M.P. (2007) An iterative algorithm for applying the theory of regions in process mining. Proceedings of the *Workshop on Formal Approaches to Business Processes and Web Services (FABPWS'07)*, pp. 36–55.
18. Bergenthum R., Desel J., Lorenz R., Mauser S. (2007) Process mining based on regions of languages. *Business Process Management: 5th International Conference (BPM 2007), Brisbane, Australia, September 24–28, 2007*, pp. 375–383. [https://doi.org/10.1007/978-3-540-75183-0\\_27](https://doi.org/10.1007/978-3-540-75183-0_27)
19. Leemans S.J.J., Fahland D., van der Aalst W.M.P. (2013) Discovering block-structured process models from event logs: A constructive approach. *Application and Theory of Petri Nets and Concurrency: 34th International Conference (PETRI NETS 2013), Milan, Italy, June 24–28, 2013*, pp. 311–329. [https://doi.org/10.1007/978-3-642-38697-8\\_17](https://doi.org/10.1007/978-3-642-38697-8_17)
20. Leemans S.J.J., Fahland D., van der Aalst W.M.P. (2015) Scalable process discovery with guarantees. *Enterprise, Business-Process and Information Systems Modeling (BPMDS EMMSAD 2015). Lecture Notes in Business Information Processing*, vol. 214. Springer, Cham, pp. 85–101. [https://doi.org/10.1007/978-3-319-19237-6\\_6](https://doi.org/10.1007/978-3-319-19237-6_6)
21. Pande S.D., Kanna R.K., Qureshi I. (2022) Natural language processing based on name entity with n-gram classifier machine learning process through ge-based hidden Markov model. *Machine Learning Applications in Engineering Education and Management*, vol. 2, no. 1, pp. 30–39.
22. Sagayam K.M., Hemanth D.J. (2019) A probabilistic model for state sequence analysis in hidden Markov model for hand gesture recognition. *Computational Intelligence*, vol. 35, no. 1, pp. 59–81. <https://doi.org/10.1111/coin.12188>
23. Srivastava R.K., Pandey D. (2022) Speech recognition using HMM and Soft Computing. *Materials Today: Proceedings*, vol. 51, pp. 1878–1883. <https://doi.org/10.1016/j.matpr.2021.10.097>
24. Du J., Wang C., Wang L. et al. (2023) Automatic block-wise genotype-phenotype association detection based on hidden Markov model. *BMC Bioinformatics*, vol. 24, article 138. <https://doi.org/10.1186/s12859-023-05265-5>
25. Rabiner L.R. (1990) A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286. <https://doi.org/10.1109/5.18626>

### About the author

#### Artem Yu. Varnukhov

Assistant, Department of Business Informatics, Ural State University of Economics, 62, 8 Marta Str., Yekaterinburg 620144, Russia;  
E-mail: [varnuhov\\_ayu@usue.ru](mailto:varnuhov_ayu@usue.ru)