

С. Ф. ЧАЛЫЙ, д-р техн. наук, проф. ХНУРЕ, Харьков;
Е. О. БОГАТОВ, асп., каф ИУС, ХНУРЕ, Харьков;
А. Ю. КАЛЬНИЦКАЯ, асс., каф. ИУС, ХНУРЕ, Харьков;

ФОРМИРОВАНИЕ НАБОРА ОГРАНИЧЕНИЙ В ЗАДАЧЕ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ НЕСТРУКТУРИРОВАННЫХ ЖУРНАЛОВ РЕГИСТРАЦИИ СОБЫТИЙ

В статье рассмотрена задача предварительной обработки журнала регистрации событий, которая решается при интеллектуальном анализе процессов. Предложен подход к формированию системы ограничений в виде набора правил при решении такой задачи. Предложена реализация подхода на основе мультиагентной системы.

Ключевые слова: мультиагентная система, интеллектуальный агент, Process mining, правила, журнальный файл.

У статті розглянута задача попередньої обробки журналу реєстрації подій, яка вирішується при інтелектуальному аналізі процесів. Запропоновано підхід до формування системи обмежень у вигляді набору правил при вирішенні такого завдання. Запропонована реалізація підходу на основі мультиагентної системи.

Ключові слова: мультиагентна система, інтелектуальний агент, Process mining, правила, журнальний файл

In this paper we consider the problem preprocessing event log, which is solved by Process mining. An approach to the formation of the system of constraints in the form of a set of rules for solving a problem. Proposed implementation of the approach based on multi-agent system.

Key words: multi-agent system, intelligent agents, Process mining, rules, log-file.

Введение. В области моделирования процессов обработки ресурсов в настоящее время существует два основных направления к построению моделей. Первое, традиционное, основано на построении моделей с использованием документации, опроса сотрудников, выявления связанных с процессами экспертных знаний и т.п. Второе, технология интеллектуального анализа процессов РМ (Process mining), которое очень интенсивно развивается в последние годы, связано с построением моделей на основе анализа журнальных файлов информационных систем, которые обеспечивают поддержку этих процессов (возможно, неформализованных). В указанных журнальных файлах регистрируются события, которые происходят в информационной системе.

Важное преимущество второго направления связано с тем фактом, что на основе журналов регистрации событий можно построить модель, отражающую реальное поведение процесса, тогда как модель на основе документации не всегда учитывает последовательности действий, отражающих скрытые неформализованные знания.

В то же время технология и методы интеллектуального анализа процессов предъявляют определенные требования к исходным данным, к исходному логу журнала. Исходные данные для построения модели процесса должны быть логически структурированы таким образом, чтобы отображать фактически выполнившиеся последовательности событий, отражающие каждый конкретный процесс обработки ресурсов (в частности, бизнес-процесс). Формирование указанных структурированных журнальных файлов осуществляется только в современных процессно-ориентированных системах в автоматическом режиме.

В то же время, существует значительное количество организаций, где используются информационные системы, не поддерживающие процессы обработки событий. Текущие действия в таких системах протоколируются с метками времени.

Иными словами, регистрационные файлы в таких системах содержат информацию, необходимую для построения процессной модели и отражающей реально выполняющиеся процессы. Однако такая информация не упорядочена в соответствии с существующими требованиями интеллектуального анализа процессов. Это и обуславливает актуальность темы статьи.

Вышеизложенное определяет актуальность проблемы предварительной обработки неструктурированных журналов регистрации событий.

Анализ исследований и публикаций. Работы в области интеллектуального анализа процессов [1–3] посвящены разработке методов и технологии построения моделей процессов обработки ресурсов на основе последовательности событий, отражающих поведение реального неформализованного процесса и зафиксированных в журнале регистрации событий, обычно имеющем следующие элементы [4, 5]: временная отметка, наименование события, объект, пользователь. Наличие указанных структурных элементов является существенным ограничением разработанных методов интеллектуального анализа процессов.

В работах [4, 5] были рассмотрены ограничения на применение технологии интеллектуального анализа процессов, а также рассмотрен метод предварительной обработки журналов регистрации событий в задачах интеллектуального анализа процессов. Данный метод предполагает формирование структурированного журнального файла из неструктурированного на основе экспертных правил, получаемых при интервьюировании экспертов в данной области [4]. Таким образом, получаемые результаты в значительной степени зависят от знаний эксперта. Однако такой подход имеет ряд недостатков: знания эксперта могут быть в значительной степени субъективными; знания предоставленные экспертом или группой экспертов могут быть противоречивыми; полученные знания могут быть неполными.

Это обуславливает актуальность разработки подходов к формированию ограничений в задаче предварительной обработки неструктурированных журналов регистрации событий.

Постановка задачи. Задача формирования набора ограничений при предварительном преобразовании журнала регистрации событий в форму, пригодную для проведения интеллектуального анализа процессов, заключается в следующем. Регистрационные файлы (регистрационные базы данных) систем коллективной работы необходимо преобразовать в структурированные журнальные файлы, аналогичные журналам регистрации событий процессно-ориентированных информационных систем.

Исходными данными задачи являются: неструктурированный регистрационный файл, содержащий в общем случае неупорядоченное описание событий с указанием метки времени. Из такого файла может быть извлечен перечень обрабатываемых объектов предметной области и их атрибутов.

Необходимо на основе анализа неструктурированного файла выявить набор ограничений, задающих возможные пути преобразования неструктурированного журнала регистрации событий в структурированный, а также обосновать возможность решения задачи предварительной обработки журнала при наличии заданных ограничений.

Основные шаги решения поставленной задачи включают в себя:

- формирование перечня объектов, с которыми оперирует процесс на основе анализа журнала регистрации событий;
- нахождение набора ограничений, отражающих возможные зависимости между объектами процесса из полученного на предыдущем шаге перечня;
- проверка возможности получения решения задачи предварительной обработки, удовлетворяющего системе полученных ограничений; в случае невозможности выполнить ограничения они должны быть изменены (ослаблены);
- разработка мультиагентной системы, реализующей преобразование журнала согласно полученной системе ограничений.

В данной статье рассматриваются следующие шаги решения поставленной задачи:

- нахождение набора ограничений, отражающих возможные зависимости между объектами процесса из полученного на предыдущем шаге перечня;
- разработка мультиагентной системы, реализующей преобразование журнала согласно полученной системе ограничений.

Нахождение набора ограничений, отражающих возможные зависимости между группами объектов процесса. Выявление перечня объектов процесса осуществляется на первых этапах метода предварительной

обработки журналов регистрации событий в задачах интеллектуального анализа процессов, путем выделения списка объектов над, которыми совершаются действия в системе.

Отметим, что объекты и группы объектов процесса могут служить основанием для выделения процедур процессов, что в свою очередь позволяет строить многоуровневое описание процесса предприятия. Для решения поставленной задачи предлагается использовать методы кластеризации, технологии интеллектуального анализа данных.

Проблема удовлетворения ограничений (CSP) состоит в нахождении решений для заданных ограничений, то есть определение допустимой области значений переменных в объектах, которые смогут удовлетворить всем ранее заданным ограничениям. Таким образом ограничения задают множества допустимых значений [7].

На сегодняшний день активно развивается направление программирования ограничений, основанная на теории проблемы покрытия/удовлетворения заданных ограничений. Согласно данной теории [7, 8] формальное представление задачи удовлетворяющих ограничений описывается представленным далее образом.

Пусть имеется, конечный алфавит переменных, который определяется, как множество переменных $K = \langle k_1, k_2, \dots, k_n \rangle$.

Пусть имеется множество объектов $T = \langle t_1, t_2, \dots, t_n \rangle$ и каждый объект может принимать одно из множества значений K , определяемых как $k_i \in t_i$, содержимое ячейки в таком случае обозначим как $k(t)$. Множество всех возможных состояний, которые могут принимать объекты представляется через $T \rightarrow K$, обозначим как k^T . Для формирования ограничений на группы объектов выделяют понятие домена/группы объектов, которое определим как T' , $T' \subset T$, таким образом, значения, принимаемые в данном подмножестве, определяются как $k(T')$. Данное обозначение $k(T')$ следует понимать как функцию формата $T' \rightarrow K$, являющуюся сужением функции $\bar{k} : T \rightarrow K$, на подмножество $T' \subset T$.

Функция, описывающая допустимость выбранного набора ограничений, определяется как $G : K^T \rightarrow \{0, 1\}$. Которая для каждого набора $\bar{k} : T \rightarrow K$ определяет его допустимость $G(\bar{k}) = 1$, или недопустимость $G(\bar{k}) = 0$. Пусть задано условие $g_{T'}(k(T'))$, которое является ограничением на существование объекта с заданной переменной.

Таки образом, формально задача имеет следующий вид:

$$G(\bar{k}) = \bigwedge_{T' \in T} g_{T'}(k(T')) = 1. \quad (1)$$

В качестве ограничения могут выступать функции, характеризующие как ограничения на принимаемые значения, так и допустимые структурные

взаимосвязи между значениями. Функции второго вида представляют типовые шаблоны взаимодействия элементов.

Каждая запись в неструктурированном журнале регистрации событий характеризуется следующим набором параметров [4]:

$$S = \langle T, A, U, O \rangle, \quad (2)$$

где S – журнальный файл;

T – временной параметр;

A – действие (событие);

U – исполнитель;

O – объект процесса, над которым совершаются действия.

Для описания шаблона посредством набора ограничений в качестве объектов T , были взяты события, записанные в журнальном файле. В качестве переменных k , которые могут содержать объекты, был выбран параметр времени. Данный выбор был связан с тем, что все объекты журнального файла синхронизированы и структурированы по временному параметру, а также следует отметить, что данный параметр позволяет выстроить последовательность наступления того или иного события при исполнении рассматриваемого процесса.

В качестве переменных, которые принимают рассматриваемые нами объекты, целесообразно использовать параметр времени наступления события. Следует отметить, что для описания структурных шаблонов достаточно представить свойства и отношения переменной одного объекта относительно другого.

Таким образом, решение поставленной задачи сводится к определению допустимого набора ограничений на значения переменных, которые характеризуют свойства объектов из рассматриваемого множества.

В основе задач моделирования процессов лежат знания о структурных шаблонах процессов. Следует отметить, что любой шаблон, применяемый при построении модели процесса, можно представить как набор ограничений.

Выделяют следующие базовые структурные шаблоны:

- последовательность;
- параллельное разбиение;
- синхронизация;
- эксклюзивный выбор;
- простое слияние.

Каждый структурный шаблон является последовательностью событий, либо подпроцессов, что в свою очередь зависит от рассматриваемого нами уровня представления процесса. Структурные шаблоны можно описать как систему из 3 объектов.

$$P = \langle O, O, O \rangle, \quad (3)$$

где P – шаблон;

O – рассматриваемый объект.

Последовательность определяет безусловный переход от одной процедуры бизнес-процесса к другой, представлена на рис. 1. Таким образом, ограничением на отнесение последовательности событий к данным шаблонам является следующее выражение:

$$k_{(t_1)} < k_{(t_2)} < k_{(t_3)}. \quad (4)$$

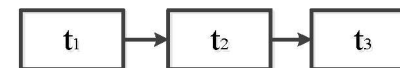


Рис. 1 – Шаблон «Последовательность»

Параллельное ветвление (AND – ветвление) – определяет разбиение единого потока управления на множество параллельных потоков, представлена на рис. 2.

Каждый параллельный поток реализуется в виде одной или нескольких последовательно выполняющихся процедур бизнес-процесса. Инициализация следующих за ветвлением событий происходит одновременно и оба события следуют за одним и тем же объектом, следовательно, данный шаблон описывается следующим набором ограничений:

$$k_{(t_1)} < k_{(t_2)} \ \& \ k_{(t_1)} < k_{(t_3)} \mid k_{(t_2)} = k_{(t_3)}. \quad (5)$$

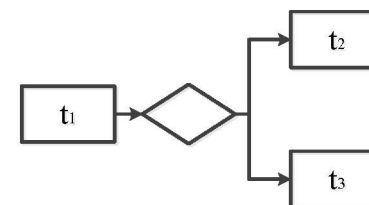


Рис. 2 – Шаблон «Параллельное ветвление»

Синхронизация (AND – объединение) – определяет синхронизируемое слияние множества параллельных процессов (последовательностей процедур) в один поток управления, причем предполагается, что каждая входная ветвь синхронизатора выполняется только один раз, представлена на рис. 3.

Инициализация предшествующих ветвлению событий происходит одновременно и оба события предшествуют последующему, следовательно, данный шаблон описывается следующим набором ограничений:

$$k_{(t_1)} > k_{(t_2)} \ \& \ k_{(t_1)} > k_{(t_3)} \mid k_{(t_1)} = k_{(t_2)}. \quad (6)$$

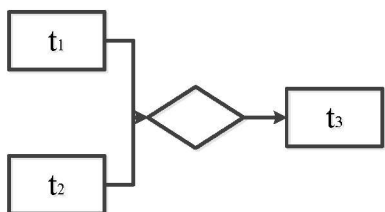


Рис. 3 – Шаблон «Синхронизация»

Эксклюзивный выбор (XOR – разделение) – определяет выбор, в зависимости от управляющих данных, одной из нескольких ветвей реализации процесса, представлена на рис. 4.

Следующие за ветвлением события t_2 , t_3 наступают в разный момент времени, каждый из объектов попеременно следуют за одним и тем же объектом t_1 ; следовательно, данный шаблон описывается следующим набором ограничений:

$$\begin{aligned}
 & (k_{(t_1)} < k_{(t_2)} \parallel k_{(t_1)} < k_{(t_3)}) \mid \\
 & | k_{(t_2)} = \text{true} \ \& \ k_{(t_3)} = \text{false} \mid | k_{(t_2)} = \text{false} \ \& \ k_{(t_3)} = \text{true} . \quad (7)
 \end{aligned}$$

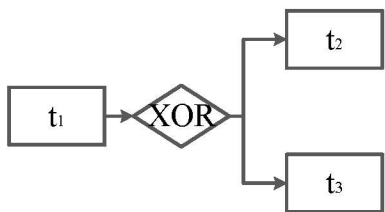


Рис. 4 – Шаблон исключающее «ИЛИ»

Простое слияние (OR – объединение) – определяет схождение нескольких альтернативных ветвей без синхронизации при условии, что ни одна из процедур, принадлежащих альтернативным ветвям, не выполнялась параллельно, представлена на рис. 5.

События, предшествующие ветвлению t_1 и t_2 наступают попеременно, оба события t_1 и t_2 предшествуют событию t_3 ; следовательно, данный шаблон описывается следующим набором ограничений.

$$\begin{aligned}
 & (k_{(t_1)} > k_{(t_3)} \parallel k_{(t_2)} > k_{(t_3)}) \mid \\
 & | k_{(t_1)} = \text{false} \ \& \ k_{(t_2)} = \text{true} \mid | k_{(t_1)} = \text{true} \ \& \ k_{(t_2)} = \text{false} . \quad (8)
 \end{aligned}$$

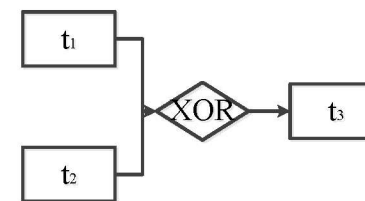


Рис. 5 – Шаблон «Простое слияние»

Для реализации представленных ограничений используется мульти-агентный подход, как рассмотрено далее. Указанный набор ограничений в форме структурных шаблонов построения модели является базовым, однако, даже данный минимальный набор ограничений, при построении модели, может вызвать противоречивость при исполнении модели. Противоречивость при построении модели может быть связана с тем, что принятие решения о отнесении последовательности из нескольких объектов к какому-либо шаблону осуществляется на основе информации о частоте повторения данной последовательности во всем анализируемом файле регистрации событий. Очевидно, что могут возникать ситуации, при которых невозможно однозначно отнести рассматриваемую последовательность к имеющимся в базе знаний шаблонам, либо в базе знаний может отсутствовать информация о данном шаблоне, для разрешения таких проблем предлагается использовать алгоритмы, которые рассмотрены в работах по формальным моделям образного мышления [7, 8].

Мультиагентный подход к реализации системы ограничений. Агент (интеллектуальный агент) определяется как автономная сущность, которая ведет наблюдение за внешней средой (получает информацию) по средствам сенсоров, оказывает на неё воздействие с помощью имеющихся механизмов и направляет свою деятельность для достижения поставленных целей [6].

Типовая структура интеллектуального агента содержит следующие элементы:

- сенсоры;
- блок обработки текущего состояния внешней среды/ состояние наблюдаемого объекта;
- множество продукционных правил (если – то);
- блок выбора действия согласно исполняемому правилу;
- механизмы агента.

Для принятия решений интеллектуальные агенты (ИА) могут использовать как алгоритм, так и базу знаний, а так же поддерживать самообучение. Для мультиагентной системы разрабатываемой, для реализации преобразования журнала на основе системы ограничений, используется база знаний. Данный выбор связан с тем, что базовые структурные шаблоны, которые предполагается хранить в базе знаний системы, могут быть дополнены другими

шаблонами моделирования процессов, либо базироваться на уже имеющихся в базе знаний шаблонах.

Интеллектуальные агенты, функционирующие в рамках разрабатываемой системы должны удовлетворять следующим требованиям:

- автономность – агенты являются частично или полностью автономными;
- локальность – каждый агент располагает лишь частичной информацией об объекте/окружении;
- децентрализация – процесс управления поиском решений распределен между ИА.

Таким образом, следует выделить группу агентов распознавания и группу ИА управления. Агенты группы распознавания по типу взаимодействия относятся к агентам групповой коммуникации.

Взаимодействие между такими агентами осуществляется следующим образом:

- каждый агент группы знает о существовании другого агента группы, а также о совершаемых им в данный момент действиях, такого рода информацию можно получить при запросе состояния другого агента;
- решение задачи зависит от результатов участия каждого из агентов, агенты по отдельности выполняют лишь части задачи, т.е. распознают в выявленных последовательностях шаблон.

Агенты управления соответствуют типу агентов-сервисов/ агентов взаимодействия. Данный тип агентов, по сути, представляет определенные функции, в которых могут нуждаться другие агенты, входящие в мультиагентную систему. К таким методам разработанной системы следует отнести: формирование списка часто встречающихся последовательностей, проверка на допустимость заданного набора правил. Под формированием часто встречающихся последовательностей подразумевается решение типовой задачи Data mining, которая позволяет выявить частоту вхождения в журнальный файл последовательностей заданной длины в определенную цепочку событий. Следует отметить, что для решения поставленной задачи – выявления шаблонов проектирования достаточно знать о последовательностях с размерностью в 3 элемента, согласно правилами формирования отношений порядка [7] и работам по формальному описанию образного мышления [8]. Общая схема разработанной мультиагентной системы представлена на рис. 6. Для принятия решений в агентной системе используются: база знаний, механизмы вывода, наблюдения (информация, поступающая с датчиков/сенсоров), механизмы / методы воздействия на внешнюю среду.

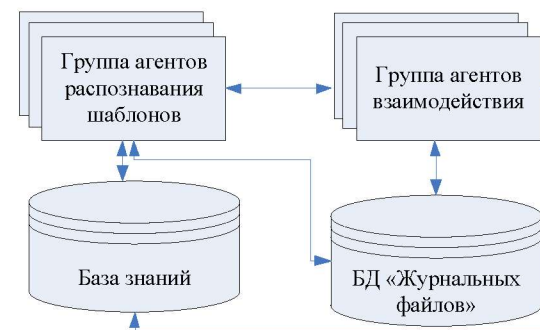


Рис. 6 – Структурная схема разработанной мультиагентной системы

Выводы. В работе выполнена постановка задачи предварительной обработки неструктурированных журналов регистрации событий на основе системы формализованных ограничений.

Представлен подход к заданию ограничений на основе шаблонов структурных взаимосвязей между объектами процессов в неструктурированных журналах регистрации событий, а также подход к предварительной обработке таких журналов на основе мультиагентных технологий. Сформированные ограничения на базовые структурные шаблоны в дальнейшем могут быть дополнены и расширены более специализированными шаблонами процесса.

Список литературы: 1. *W. M. P. van der Aalst. Process-Aware Information Systems: Lessons to be Learned from Process Mining / W. M. P. van der Aalst // Department of Mathematics and Computer Science Eindhoven University of Technology. – Eindhoven, 2000. – 26 p.* 2. *W. M. P. van der Aalst. Business Process Management Demystified : A Tutorial on Models, Systems and Standards for Workflow Management / W. M. P. van der Aalst // Department of Mathematics and Computer Science Eindhoven University of Technology. – Eindhoven, 2004. – 68 p.* 3. *W. M. P. van der Aalst. Process Mining : Discovery, Conformance and Enhancement of Business Processes / W. M. P. van der Aalst. – N.–Y. : Springer Verlag. – 2011. – 370.* 4. *Левыкин В. М. Метод предварительной обработки журналов регистрации событий в задачах интеллектуального анализа процессов / В. М. Левыкин, С. Ф. Чалый, Е. О. Богатов, М. И. Дикусар // Сборник статей № 7 (88). – К., 2011. – ДП «ЦНДИНУ». – С. 203–207.* 5. *Чалый С. Ф. Предварительное преобразование журнальных файлов в задачах Process mining / С. Ф. Чалый, Е. О. Богатов // Збірка матеріалів другої міжнародної науково-технічної конференції “Сучасні напрями розвитку інформаційно- комунікаційних технологій та засобів управління”. – Київ, 2012. – ДП «ЦНДИНУ», 2011. – 51 с.* 6. *Stuart Russell. Artificial Intelligence: A Modern Approach (3rd Edition) [Hardcover] / Stuart Russell, Peter Norvig // Publication Date : December 11, 2009 | ISBN-10 : 0136042597 | ISBN-13 : 978-0136042594 | Edition : 3.* 7. *Schlesinger M. I. Solution to Structural Recognition (MAX,+)-problems by their Equivalent Transformations. Part 1. In Russian: Решение (MAX,+)-задач структурного распознавания с помощью их эквивалентных преобразований. Часть 1. Control Systems and Computers, 2007(1), Kiev. – P. 3–15.* 8. *Schlesinger M. I., Giginyak V. V. Solution to Structural Recognition (MAX,+)-problems by their Equivalent Transformations. Part 2. In Russian : Решение (MAX,+)-задач структурного распознавания при помощи их эквивалентных преобразований. Часть 2.*

Надійшла до редколегії 06.12.2012