РАЗДЕЛ I. ТЕХНИЧЕСКИЕ НАУКИ

Балтрашевич В.Э.

К вопросу получения и представления знаний о процессах

*Санкт-Петербургский Государственный электротехнический университет
(Россия, Санкт-Петербург)*

doi: 10.18411/sr-10-02-2020-01

idsp: sciencerussia-10-02-2020-01

Аннотация

На примере простейшей предметной области демонстрируется процесс извлечения знаний о процессах. Подчеркивается необходимость различать процесс извлечения знаний (ПИЗ) и исследуемый процесс (ИП). Отмечаются особенности обоих процессов. Для анализа ИП используется системный анализ, который лежит в основе ПИЗ. Анализируются различные формы представления знаний об ИП. Рассматриваются различные программные средства, поддерживающие ПИЗ об ИП.

Ключевые слова: знания о процессах, системный анализ, формы представления знаний, process mining, цепи Маркова

Abstract

The process of extracting knowledge about processes is demonstrated using the simplest domain as an example. The need to distinguish between the process of extracting knowledge (PIP) and the process under study (IP) is emphasized. The features of both processes are noted. For the analysis of IP, a system analysis is used, which is the basis of the PIP. Various forms of representing knowledge about IP are analyzed. Various software tools supporting PIP about IP are considered.

Keywords: knowledge of processes, system analysis, forms of knowledge representation, process mining, Markov chains

В настоящее время особое внимание уделяется извлечению знаний о процессах (process mining) [1]. Но в большинстве работ исследования проводятся для бизнес-процессов, что затрудняет выделение этапов ПИЗ. Поэтому представляется целесообразным взять в качестве предметной области простую область, например, область аналого-цифрового преобразования (ацп-я). Это позволит показать process mining (получение знаний о процессах) на простом примере, не отвлекаясь на тонкости предметной области, а акцентируя внимание на изучении (и в дальнейшем - формализации) процесса получения знаний об изучаемом процессе. При анализе ИП надо различать процесс и генератор (источник) процесса. Так аналого-цифровой преобразователь (АЦП) при воздействии входного сигнала является генератором процесса, на выходе которого имеем исследуемый процесс, состояниями которого являются числовые эквиваленты и ответы схемы сравнения (в которых косвенно учитывается входной сигнал).

Использование получения и представления знаний об ИП позволяет:

1. *Получить описание ИП на понятном (для специалистов анализируемой предметной области) языке.*
2. *Однотипно проанализировать работу генераторов и ИП в различных режимах.*
3. *Выявить используемые шкалы (единицы измерения).*

4. Проанализировать различные формы представления знаний о состояниях ИП.
5. Выявить базовые действия, входящие в ИП.
6. Легко корректировать полученные знания.
7. Использование разных форм представления знаний на разных этапах системного анализа.
8. Наглядно представить граф ИП в графическом виде, что позволит легко определить полноту базы знаний, понять ИП и его модернизировать.
9. Улучшить характеристик существующих систем за счёт синтеза обновлённых систем.

Чтобы получить знания о процессе его надо подробно исследовать, т.е. провести системный анализ [9]. Системный анализ включает в себя использование строгих математических методов и эвристик по их совместному использованию, для реализации которых часто используются экспертные системы (ЭС). Таким образом, системный анализ лежит в основе process mining. При этом надо выявить и обособить знания и о ПИЗ и об ИП.

Раньше пытались создавать средства (средства ПИЗ) для однотипного описания различных алгоритмов ацп_я (ИП). Выявлялись базовые действия ИП и разрабатывались языки описания функционирования (ЯОФ) генераторов. При разработке ЯОФ был использован аппарат регулярных выражений и метод рекурсивного спуска [7]. Но такое представление знаний было неудобно из-за 1. требования знания программирования, пусть и на более высоком уровне; 2. перекомпиляции программы на ЯОФ при изменении ИП; 3. трудностей при создании новых генераторов процессов; 4. трудностей при проверке полноты базы знаний.

Очевидно, что ЯОФ являются средствами ПИЗ. Как известно, весьма полезным является представление описание процесса с помощью графов. Для ЯОФ граф представлялся в виде блок-схемы, элементами которой были базовые действия, связанные между собой классическими процессами программирования (линейный, ветвящийся, циклический). Блок-схема является графом базы знаний о работе генератора процесса. Т.е. в этих разработках **описание генератора было первичным, а порождаемый процесс был вторичным**. Действия, конечно, использовали и цифровой код и ответы схемы сравнения, но в скрытом (от пользователя) виде. Как показала практика, полезность такого представления знаний для анализа существующих и разработки новых генераторов (а тем более самих процессов) – затруднительна.

При подходе, используемом в process mining, **первичным является описание процесса, а вторичным – описание генератора**. Работа process mining базируется на анализе экземпляров процессов. Блок-схема является графом базы знаний о работе генератора процесса. Графом процесса является граф переходов состояний в зависимости от ответов схемы сравнения [4].

Рассмотрим наш генератор (источник) исследуемого процесса как черный ящик. Берём какой-либо сигнал, например, синусоидальный, подаём на вход ящика. Ящик работает и создаёт экземпляры процесса. Получаем журнал событий (англ. event logs), в котором фиксируются возможные экземпляры процессов. Заметим, что выявление экземпляров - это этапы ПИЗ. Под событием понимаем пару: образцовый уровень, ответ схемы сравнения. Пытаемся построить граф, отражающий формальное описание процесса на основе журнала. Нарисовать полный граф ИП по нескольким экземплярам процесса не удастся.

Для примера, строим частичный граф поразрядного трех разрядного преобразователя (в двоичном коде) при обработке конкретного сигнала, равного 4.5. Видим 100 (110 101)... 100, т.е. начинаем и кончаем в одном состоянии. Как их различить? Видно, что нужно использовать два уровня по времени. (000 100) (100 110) (110 101) (101 100), т.е. начинаем и кончаем в разных состояниях и к каждому из них могут быть применены свои действия.

Графическое представление облегчает анализ процесса: 1. видно, что база знаний (граф) построенная по частичным процессам не является полной, т.е. отсутствуют все возможные переходы; 2. видно, что не всегда достаточно одного образцового уровня, а надо два или более; 3. видны выходы за границы диапазона.

Был предложен подход, базирующийся на экземплярах процессов, которые предлагалось рассматривать как цепи Маркова [2 – 6, 8, 10, 11]. Появилась возможность однотипно описывать различные процессы (алгоритмы) ацп-я и возможность разработки и создания широкого круга инструментальных средств по их анализу. Под инструментальными средствами понимаем средства, которые остаются неизменными и позволяют только менять описания исследуемых процессов.

Выбираем классические (хорошо отлаженные) алгоритмы. Берем их программную (процедурную) реализацию и на ней проверяем правильность разрабатываемого подхода. После чего можем разрабатывать новые структуры генераторов, порождающих исследуемые процессы и их анализировать, извлекая знания и их используя.

Предложен способ и его реализация построения полного графа (по существующей процедуре преобразования) [8]. Т.е. разработана инструментальная система построения полного графа (средство ПИЗ). Её целесообразно снабдить средствами редактирования и пересортировки графа ИП. Разработанная программа (инструментальная система) генерации всех состояний и их переходов, строит (это этапы ПИЗ) таблицу состояний и их нумерации, и таблицу возможных переходов. Таблицы – это знания об ИП. Построение таблиц – это этапы ПИЗ.

Можем строить графы различных процессов. Используя эти средства можно разрабатывать новые процессы преобразования – ИП. Эксперт рисует граф процесса из своих представлений о повышении эффективности. Для проверки работоспособности предложенного алгоритма (для детального пошагового анализа) используется имитационное моделирование, как составная часть системного анализа.

Существуют два способа реализации (имитации) работы графа ИП.

1. По матрице переходов и по матрице состояний строим правила продукции.

2. Эксперт должен проанализировать переходы между состояниями в зависимости от ответов схемы сравнения и подобрать базовые действия для их реализации. Чтобы облегчить эксперту (не специалисту в программировании) описание процессов, предлагается использовать инструментальные экспертные системы, в которых базовые действия задаются с помощью глубинных знаний, а сам процесс описывается на языке продукции, близком к естественному.

Проводим имитационное моделирование при разных входных сигналах (соответствующих наиболее часто встречающимся распределениям). Анализируем протекание ИП. Поучаем характеристики ИП, которые являются декларативными знаниями об ИП (этап ПИЗ). Проверяем шкалы, корректируем шкалы (это этапы ПИЗ). А результаты – характеристики ИП.

Инструментальная система (программа) имитационного моделирования (этап ПИЗ). Исполняющей системой в этом случае является инструментальная ЭС.

Системный анализ требует проведения двухэтапного анализа, т.е. наряду с имитационным моделированием требует проведения статистического или вероятностного моделирования. Введённое ранее представление знаний о процессе с помощью таблиц номеров состояний и таблицы переходов состояний позволяет использовать аппарат цепей Маркова для проведения вероятностного моделирования изучаемого процесса. Причем заметим – аналитического анализа.

Разработана инструментальная система вероятностного моделирования (этап ПИЗ). Вероятностный анализ – это вариант интегрального подхода к анализу. Программа вероятностного моделирования позволяет выявить возможности повышения точности преобразования.

Следует отметить, что представление процесса и в виде таблиц и в виде графа и в виде описания на языке эксперта являются различными формами представления знаний о процессе (ИП). Возможно их взаимное преобразование, т.е. компиляция знаний.

Заметим, что везде в ходе изложения акцентировалось внимание, к каким процессам относятся те или иные действия. (ПИЗ, сам ИП, генератор ИП).

Выше были выявлены этапы (составляющие) ПИЗ и этапы (составляющие) ИП. Они приведены в таблице 1.

Таблица 1.

Составляющие ПИЗ и ИП

ПИЗ	ИП
Выявление шкал	Коррекция шкал
Выявление базовых действий.	Набор базовых действий ИП
Разработка и использование ЯОФ	Описание ИП на ЯОФ
Построение графа ИП (блок-схема)	Блок-схема ИП
Выявление экземпляров ИП	Набор экземпляров ИП
Инструментарий для выбора состояний	Набор (множество) состояний
Построение частичного графа состояний	Граф состояний (не полный)
Инструментальная система имитационного моделирования	Имитационное моделирование ИП
Технология (инструментальная система) построения всех состояний и их связи	Таблица номер-состояние и таблица переходов ИП
Графический (инструментальная система) редактор построения графа состояний	Полный граф состояний и их переходов
Построение матрицы переходных вероятностей	Матрица переходных вероятностей ИП
Инструментальная система (программа) вероятностного моделирования	Вероятностный анализ и вероятностное моделирование ИП
Разработка форм представления знаний	Формы представления знаний об ИП
Инструментальные системы (программы) взаимного преобразования (компиляции) форм представления знаний	Различные формы представления знаний

Таким образом, в работе рассматривается опыт использования системного анализа для получения знаний о процессе адм_я. При этом рассматривались следующие задачи:

- построение оптимальной шкалы квантования, при этом с одной стороны используются строгие математические методы адаптивного обучения, обеспечивающие достижение экстремума некоторого функционала определяемого задаваемой метрикой, с другой стороны, необходимо провести коррекцию шкалы (т.е. оптимизацию по эвристическому критерию) на основании рекомендаций эксперта, стремящегося упростить техническую реализацию шкалы. Особое внимание уделяется вопросам, связанным с представлением шкал и с представлением знаний, позволяющих корректировать параметры

шкалы и организовывать связь с пользователем. Решение задачи построения оптимальных шкал демонстрирует процесс формирования значений атрибутов при формализации знаний в трудно формализуемых предметных областях, т.е. один из аспектов обучения экспертных систем;

- в соответствии с принципом двухэтапной оптимизации системного анализа, генераторы ИП должны быть проанализированы в двух режимах: при воздействии детерминированного сигнала и при воздействии детерминированного сигнала с наложенным на него некоторым шумом, а также при воздействии случайного процесса. Это приводит к необходимости реализации имитационного и вероятностного моделирования;
- описание структур аналого-цифровых преобразователей (генераторов ИП). Обсуждается представление базовых элементов, позволяющих описывать генераторы ИП с помощью многократных и многомерных цепей Маркова;
- описание ИП с помощью различных форм представления знаний, при этом рассматриваются задачи двух классов: формализация знаний эксперта с помощью специальных редакторов и автоматическое формирование знаний о процессе преобразования уже готового АЦП. Реализация редакторов базируется на объектах, описывающих элементы АЦП, и связанных с ними действиях, и позволяет описывать процесс преобразования на языке близком к естественному языку;
- компиляция знаний о процессе преобразования, описанного с помощью одной из форм представления знаний и отлаженного с помощью имитационного и вероятностного моделирования;
- имитационное и вероятностное моделирование процесса аналого-цифрового преобразования (ИП) заданных структур АЦП (генераторов ИП).

1. Анализ данных и процессов: учеб. пособие/А.А. Барсегян, М.С. Куприянов, И.И. Холод и др. -3-е изд., перераб. И доп. – СПб.: БХВ-Петербург, 2009. – 512с.
2. Балтрашевич В.Э. Использование цепей Маркова для вероятностного анализа аналого-цифровых преобразователей. "Изв. ВУЗов. Электромеханика", 1977, №11. с.1195-1200.
3. Балтрашевич В.Э. Вероятностный анализ следящих АЦП. «Известия ЛЭТИ», 1981. вып.291. с.11-16.
4. Балтрашевич В.Э. Моделирование процесса аналого-цифрового преобразования. «Известия ЛЭТИ», 1981. вып.293. с.78-82.
5. Балтрашевич В.Э. Анализ процесса преобразования в многоуровневых аналого-цифровых преобразователях. "Изв. ВУЗов. Электромеханика", 1983, №5. с.56-59.
6. Балтрашевич В.Э. Диалоговая САПР следящих АЦП. //Автоматизированное проектирование в радиоэлектронике и приборостроении: Межвуз. сб. науч. тр. /ленингр. электротехн. ин-т им. В.И.Ульянова (Ленина). – Л., 1987. – с.84-87.
7. Балтрашевич В.Э., Тихомирова Е.А. Вероятностное моделирование следящих АЦП. №50860000992. Инф. бюллетень «Алгоритмы и программы» ВНИИЦентра. – 1987. – вып.5.
8. Балтрашевич В.Э., Интеллектуальная поддержка системного анализа информационных процессов // Известия ЛЭТИ.-СПб.: Изд-во СПб ГЭТУ<ЛЭТИ>,2002.-С. 27-31.
9. Балтрашевич В.Э. Представление знаний о процессах с помощью цепей Маркова. //Изв. СПбГЭТУ «ЛЭТИ». 2016. №9. С. 22 – 24.
10. Балтрашевич В.Э. Анализ процесса аналого-цифрового преобразования с помощью цепей Маркова. Свидетельство о государственной регистрации программ для ЭВМ № 2017611042. 2017.