

УДК 681.3

В. Э. Балтрашевич

Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Представление знаний о процессах с помощью цепей Маркова

Описывается программный комплекс системного анализа аналого-цифрового преобразования телеметрических сигналов с интеллектуальной поддержкой. Показывается, что использование цепей Маркова для представления знаний о процессе преобразования позволяет однотипно описывать различные структуры преобразователей и облегчает проведение системного анализа, состоящего из имитационного и вероятностного моделирования. Обсуждаются вспомогательные программные средства поддержки системного анализа.

Знания, цепи Маркова, Process Mining

В настоящее время при анализе телеметрической информации широкое распространение получили методы Data Mining. В последнее время появились методы более сложной обработки, связанные с анализом процессов Process Mining [1]. В современной литературе о Process Mining анализируются сложные процессы в больших системах (учреждения, заводы и т. п.). Встает задача расширения числа пользователей этих методов. Хотелось бы при обучении этой теме использовать более простую предметную область, на примере которой можно было бы большее внимание уделить именно извлечению знаний о процессе (анализу процесса), а не изучению тонкостей одной из предметных областей. В статье рассматриваются 2 вида процессов: анализируемый процесс предметной области (АППО) и процесс анализа АППО, который сводится к системному анализу. В связи с этим представляется целесообразным на примере несложной предметной области провести анализ процессов, происходящих в ней, и уделить главное внимание процессу анализа процессов предметной области. В качестве такой предметной области в статье используется область аналого-цифрового преобразования (АЦП) телеметрической информации, рассматриваемая на примере программного комплекса системного анализа различных структур АЦП.

Существующие следящие АЦП характеризуются переменным значением шага квантования, зависящим от скорости изменения входного сигнала. Процесс формирования текущего образцового уровня таких АЦП может быть представлен в виде

$$q[i] = c_1[i]q[i-1] + c_2[i]q[i-2] + c_3[i]q[i-3], \quad (1)$$

где $q[i]$ – значение образцового уровня в i -й момент времени; $c_1[i]$, $c_2[i]$, $c_3[i]$ – числовые последовательности, зависящие от алгоритма преобразования и от входного сигнала.

Так как значения коэффициентов c_1 , c_2 , c_3 зависят от входного сигнала, то в общем случае их можно рассматривать как случайные. Учитывая, что число возможных образцовых уровней q конечно, можно считать, что выражение (1) задает дискретный случайный процесс с конечным множеством состояний.

Если же ввести вектор состояния, обозначив $q_1[i] = q[i-1]$, $q_2[i] = q[i-2]$, то выражение (1) можно записать в виде

$$\begin{bmatrix} q[i] \\ q_1[i] \\ q_2[i] \end{bmatrix} = \begin{bmatrix} c_1[i] & c_2[i] & c_3[i] \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} q[i-1] \\ q_1[i-1] \\ q_2[i-1] \end{bmatrix}. \quad (2)$$

Таким образом, если под состоянием понимать 3 образцовых уровня, то получаем зависимость только от одного предыдущего состояния, и, следовательно, процесс преобразования может описываться цепью Маркова. При этом если все 3 уровня формируются последовательно во времени, то процесс преобразования (1) описывается трехкратной цепью Маркова; если все 3 образцовых уровня формируются одновременно, то процесс преобразования (2) описывается трехмерной цепью Маркова. Так как в АЦП наряду с большим числом возможных состояний переход из

одного состояния возможен в небольшое число соседних состояний, то матрица переходных вероятностей будет сильно разреженной, т. е. будет состоять почти из одних нулей [2]. В связи с этим при моделировании для экономии памяти ЦВМ целесообразно использовать связанное представление матрицы либо ввести матрицу с четным числом столбцов и числом строк, равным числу состояний, в нечетные столбцы которой заносятся номера ненулевых состояний, а в четные – соответствующие вероятности перехода. Определять возможные переходы целесообразно с помощью имитационного моделирования. В работах [2], [3] показана возможность анализа процесса аналого-цифрового преобразования с помощью цепей Маркова. Наряду с простыми цепями Маркова рассматриваются многократные, многомерные и смешанные цепи Маркова. Для анализа преобразователей, как использующих запоминание на входе, так и не использующих его, применяют вложенные цепи Маркова, а для анализа процесса преобразования с переменным временем преобразования используют полумарковские процессы.

Проведенный анализ показал, что аппарат цепей Маркова удобно использовать для представления знаний о процессах исследуемой предметной области, что позволяет однотипно описывать целый класс различных устройств.

Таким образом, представление знаний о процессе (АППО) можно представить с помощью:

- описания текущего состояния, включающего в себя несколько эталонных уровней либо по времени, либо в пространстве, либо смешанных;
- матрицы возможных состояний и их номеров;
- перечня возможных входных сигналов для текущего состояния;
- матрицы переходов из состояния в состояние при воздействии разных входных сигналов;
- начального состояния преобразователя, определяющего строку начального распределения вероятностей при вероятностном анализе;
- перечня продукций, описывающих процесс преобразования.

В качестве исполняющих систем при реализации этих знаний должны использоваться система матричных операций (для реализации математического аппарата цепей Маркова) и инструментальная экспертная система – для реализации системы продукций.

Рассмотрим особенности формирования этих форм представления знаний.

Как в классическом Process Mining, будем предполагать, что имеется некая модель исследуемого АЦП, записанная на каком-то языке программирования. Это требует от анализатора знания языков программирования, что затрудняет его работу. На ее вход подаются различные входные сигналы и в процессе преобразования записываются в журнал все возможные появляющиеся состояния и переходы между ними. Фактически это использование имитационного моделирования, в результате которого нет гарантии, что будут получены все состояния и все переходы. Другими словами, если рассматривать граф возможных состояний и переходов между ними как базу знаний о процессе преобразования, то эта база знаний может быть неполной. Чтобы получить полную базу знаний, надо модифицировать модель так, чтобы можно было задать конкретное начальное состояние преобразователя и конкретное значение входного сигнала и получить значения недостающих переходов.

После выявления всех возможных состояний и переходов между ними необходимо с помощью методов Data Mining выявить метаданные. Для выявления метаданных и проверки полноты базы знаний целесообразно использовать графическое представление состояний и переходов между ними. Получение метаданных позволяет описать процесс преобразования различных АЦП однообразным способом, при котором по текущему состоянию и заданному входному сигналу определяется новое состояние преобразователя.

Системный анализ предполагает необходимость анализа и при имитационном, и при вероятностном моделировании. При *имитационном моделировании* происходит анализ при каждом значении входного сигнала, поэтому целесообразным представляется описание процесса преобразования с помощью системы продукций, которое легко получить, имея описание состояний и матрицу переходов состояний при воздействии входного сигнала. Для реализации системы продукций нужна исполняющая система, в качестве которой можно использовать инструментальную экспертную систему.

На основании предложенного подхода реализована система системного анализа, позволяющая однотипно описывать различные АЦП, проводить их анализ и разрабатывать (проектировать) новые структуры.

При *вероятностном моделировании* в простейшем случае задается конкретное значение сиг-

нала и на него накладывается нормальный шум с заданными параметрами. По этим значениям строится матрица переходных вероятностей, имея которую и строку начального распределения, с помощью аппарата цепей Маркова можно полностью проанализировать процесс преобразования.

Предложенный подход позволяет не только анализировать существующие устройства, но и создавать новые. Пользователь может нарисовать граф состояний и переходов, т. е. разработать новый алгоритм и с помощью предлагаемого подхода проанализировать работоспособность своего устройства как при имитационном, так и при вероятностном моделировании.

Таким образом, программный комплекс системного анализа состоит из следующих основных систем:

- определения состояния;
- выявления всех состояний и их нумерации;
- выявления метаданных;
- построения графа описания процесса;

– построения системы продукций по графу предметной области;

– имитационного моделирования;

– ведения журналов, учитывающих отдельные реализации возможных процессов;

– построения матрицы переходов и матрицы переходных вероятностей для заданного значения сигнала и характеристик шума;

– вероятностного моделирования.

Предлагаемый подход предполагает значительные затраты памяти, но позволяет проводить аналитический анализ (*с количественными оценками*) изучаемых процессов.

Интересно отметить, что рассматриваемая предметная область позволяет продемонстрировать большинство методов Data Mining. Кластеризация лежит в основе построения шкал АЦП, классификация лежит в основе алгоритма преобразования, т. е. проведенный анализ можно рассматривать как анализ процесса классификации.

СПИСОК ЛИТЕРАТУРЫ

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.
2. Балтрашевич В. Э. Использование цепей Маркова для вероятностного анализа АЦП // Электромеханика. 1977. № 11. С. 1195–1199.

3. Балтрашевич В. Э. Анализ процесса преобразования в многоуровневых аналого-цифровых преобразователях // Электромеханика. 1983. № 5. С. 56–59.

V. E. Baltrashevich

Saint Petersburg Electrotechnical University «LETI»

PRESENTATION OF KNOWLEDGE ABOUT PROCESSES WITH MARKOV CHAINS

It describes software complex system analysis of analog-to-digital conversion of telemetry signals from the intellectual support. It is shown that the use of Markov chains to represent knowledge about the process of conversion allows the same type of structure to describe a variety of transducers and facilitates the analysis of the system, consisting of a simulation and probabilistic modeling. We discuss supporting software support system analysis.

The knowledge, of Markov Chain, Process Mining