

# Метод трансформации данных для моделирования технических объектов

Н. А. Жукова, А. И. Водяхо  
СПбГЭТУ «ЛЭТИ»  
nazhukova@mail.ru

Д. К. Левоневский  
Санкт-Петербургский институт информатики и  
автоматизации РАН  
dlewonewski.8781@gmail.com

А. С. Симоненко  
Университет ИТМО  
alena\_simonenko2014@mail.ru

**Аннотация.** Предлагаемый метод позволяет строить адаптивные процессы обработки данных, поступающих от технических объектов. Адаптивная обработка осуществляется, исходя из содержания информационных потоков, при этом учитываются условия сбора, обработки и предполагаемого использования результатов. Метод основан на построении моделей трансформаций, позволяющих определять целесообразные способы обработки данных об объектах. Эффективность метода подтверждается результатами экспериментальных исследований.

**Ключевые слова:** трансформация данных; адаптивная обработка; методы синтеза

## I. ВВЕДЕНИЕ

Современный уровень развития технологической инфраструктуры предметных областей характеризуется наличием большого числа технических объектов и систем объектов. Основным источником данных об этих объектах являются поступающие от них потоки данных [1]. Потоки содержат информационные сообщения. Как правило, это результаты измерений параметров этих объектов.

Обработка данных осуществляется специализированными информационными системами [2], разработанными для решения прикладных задач в предметных областях. Эти системы предъявляют жесткие требования к составу и структуре поступающих потоков. При изменении характеристик потоков возможностей систем оказывается недостаточно для их обработки. В связи с этим актуальной является задача формирования процессов обработки данных, структура и логика которых строится и перестраивается в зависимости от условий обработки.

К условиям, определяющим формируемые процессы обработки, относятся: контент информационных потоков, контекст, в котором осуществляется обработка, цель обработки.

Под контентом понимается смысловое содержание информационных потоков, поступающих от объектов.

Контент характеризуется множеством свойств. К основным свойствам относятся содержательность, структурированность, качество, согласованность, актуальность, полнота и другие. Обработка контента осуществляется в контексте. Контекст определяется совокупностью факторов, оказывающих прямое или косвенное влияние на процессы обработки. В составе влияющих факторов могут рассматриваться параметры геоинформационной среды, характеристики средств сбора и передачи данных, информация, получаемая в результате обработки статистических данных и другая [3]. В рамках контекста также учитываются особенности предметной области.

Целью обработки является формирование представлений исходных данных в соответствии с требованиями ожидаемых потребителей. Такие представления могут восприниматься как модели, отражающие отдельные аспекты состояния или поведения объектов реального мира.

Процессы обработки контента в соответствии с контекстом, формируемые, исходя из целей обработки, будем называть адаптивными процессами.

Во втором разделе рассматриваются основные особенности адаптивных процессов. В третьем разделе предлагаются модели и метод трансформации, позволяющие строить такие процессы. В заключении рассматриваются перспективные варианты дальнейшего развития метода.

## II. ОСОБЕННОСТИ АДАПТИВНЫХ ПРОЦЕССОВ ОБРАБОТКИ

Процессы адаптивной обработки характеризуются набором параметров. Такими параметрами являются:

1. Уровни общности решаемых задач обработки. Задачи могут решаться на четырех уровнях: результатов измерений параметров, объектов, ситуаций, поддержки принятия решений [3].

2. Уровень абстракции описания данных. По поступающим данным могут строиться модели отдельных узлов или агрегатов, модели подсистем объектов, модели объектов в целом, а также групп взаимосвязанных объектов.
3. Уровень работы со знаниями. Выделяется три уровня: работа со скрытым знанием, с явным неформализованным знанием, с формализованным знанием. Формализованное знание может быть частично структурированным, структурированным или систематизированным.
4. Уровень сложности пространства обработки данных. Возможны три варианта. Первый вариант предусматривает обработку поступающих сырых данных, второй – формирование признаков пространства и выполнение преобразований над данными в новом пространстве. Третий вариант предполагает представление данных в соответствии с требованиями стандартов и работу в пространстве открытых связанных данных.

При формировании процессов обработки также учитывается тип обрабатываемой информации. Типы информации могут определяться с различных позиций. Исходя из динамических свойств информация может быть отнесена к статической или динамической. В зависимости от времени поступления она может считаться оперативной или исторической.

Изменяющиеся условия обработки данных и ее особенности определяют необходимость управлять построением и перестройкой адаптивных процессов. При этом управляющие воздействия напрямую зависят от уже полученных результатов анализа данных.

При адаптивной обработке поставленная цель должна быть достижима за конечное число шагов. В противном случае должна быть обоснована невозможность достижения цели и определен состав необходимых для ее достижения дополнительных данных.

Для построения процессов обработки и управления этими процессами предлагается использовать аппарат трансформаций [4]. Процесс обработки представляет собой последовательность целенаправленных управляемых трансформаций исходных данных, обеспечивающих формирование представлений данных в соответствии с предъявляемыми к ним требованиями. Под трансформацией данных понимается некоторое преобразование или совокупность преобразований, которые могут быть выполнены над данными.

Преобразования представляют собой функции, переводящие некоторое представление данных в другое [5]. В качестве функций могут выступать различные методы анализа, в частности, статистические, интеллектуальные, методы машинного обучения и другие. Для оценки этих методов задаются функционалы.

Для выбора управляющих воздействий необходимо решить оптимизационную задачу. Целевую функцию определяет ожидаемый результат обработки, а

ограничения – свойства контента и контекста. Для снижения сложности построения адаптивных процессов назначаются стратегии обработки и множество правил управления для каждой стратегии. Стратегия определяет способ систематизации трансформаций, выполняемых над данными, за счет определения множеств допустимых правил управления обработкой. Допустимые множества перестраиваются при переходах от одного представления данных к другому.

Трансформации могут выполняться последовательно или параллельно. Формируемое множество трансформаций является реализацией стратегии на множестве управляющих правил и множестве функционалов. В ходе обработки уровни обработки могут изменяться. В результате адаптивные процессы представляются в виде многоуровневой структуры, элементы которой определяют состав трансформаций. Между элементами устанавливаются направленные связи.

В следующих разделах предлагаются формализованная модель трансформаций и метод, позволяющий строить процессы трансформаций, способные обеспечить адаптивную обработку данных о технических объектах.

### III. МОДЕЛИ ТРАНСФОРМАЦИЙ

В качестве управляемых объектов при адаптивной обработке данных выступают процессы их обработки. Определим исходное пространство для описания этих процессов. Для этого введем ряд обозначений.

Обозначим за  $K = \{\xi\}$  класс управляемых процессов обработки контента, за  $\Pi = \{\zeta\}$  – класс частично управляемых случайных процессов, определяемых контекстом обработки. Причем процессы  $\zeta$  оказывают непосредственное влияние на процессы  $\xi$ . За  $\Phi = \{\varphi\}$  обозначим класс функционалов, определяемых на множестве преобразований, применяемых при обработке контента в контексте.

Зададим  $(X, \mathcal{M})$  – измеримое фазовое пространство процесса обработки контента в контексте, где  $\mathcal{M}$  –  $\sigma$ -алгебра измеримых множеств из  $X$ ,  $(Y, \mathcal{A})$  – пространство управлений, где  $\mathcal{A}$  –  $\sigma$ -алгебра измеримых множеств из  $Y$  и  $(Q, \Omega)$  – пространство процессов обработки, где  $\Omega$  –  $\sigma$ -алгебра измеримых множеств из  $Q$ . На пространстве элементарных процессов может быть определена вероятностная мера  $P$ . В этом случае пространство  $(Q, \Omega, P)$  представляет собой вероятностное пространство.

Представим управляемый процесс обработки контента как измеримое условное отображение вида  $\xi: Q \xrightarrow{\varepsilon} X$ . Функционалы, применяемые при обработке, определим как измеримые отображения  $X' \times Y'$  на числовую прямую, где  $t$  – время.

Для того, чтобы выделить из бесконечного множества элементов, содержащихся в классе процессов  $K \times \Phi \times \Pi$ , какой-либо один процесс, назовем способ выбора управления  $y_t \in Y$  в каждый момент времени  $t$ . Правило управления процессом обработки представим как вероятностное отображение вида  $F: X^l \times Y^{l-1} \rightarrow Y$ , где  $l$  – глубина памяти, то есть вычисление управления по правилу в смежные моменты времени  $t-l$  и  $t$  основано соответственно на предыстории  $(x_{t-l}, \dots, x_{t-1}; y_{t-l}, \dots, y_{t-2})$  и  $(x_{t-l}, \dots, x_{t-1}, x_t; y_{t-l}, \dots, y_{t-2}, y_{t-1})$ . Процессы, формируемые по правилам, являются управляемыми.

Адаптивные процессы определим как процессы, для которых предусматривается построение и использование множества управляющих правил. Определим модель трансформаций, которая позволяет строить такие процессы:

$$Y = [\tilde{U}, T_{\zeta_t}], t < \tau. \quad (1)$$

Формируемые на основе этой модели процессы обработки приводят к цели обработки в условиях  $(\xi, \zeta, \varphi)$  за конечное время  $\tau$ . Здесь  $\tilde{U} = \cup U_F$ , где  $U_F = (X, F, Y)$  – элементарная управляющая система. Элементарная система  $U_F$  управляет процессом обработки, если после действия  $y_{t-1}$  процесс принимает значение  $x_t$ . В тот же момент управляемый случайный процесс  $\xi_t$  попадает на вход управляющей системы  $U_F$  и вызывает, согласно правилу  $F$ , очередное действие управляющей системы  $y_t$ . Отметим, что процесс обработки состоит из множества связанных элементарных управляющих систем.  $T_{\zeta_t}$  – параметрическое семейство трансформаций вида  $T: D \rightarrow D, X \rightarrow X, Y \rightarrow Y$ , где  $D \subseteq D_\infty = \cup_i D_i$  – множество правил управления,  $D_i$  – совокупность всех условных распределений на  $Y$  вида  $F(N | \bar{x}_{[l]}, \bar{y}_{[0:l-1]})$ ;  $\zeta_t = \Psi(\xi, \zeta)$  – статистика процесса  $\xi$  в контексте  $\zeta$ , являющаяся измеримым отображением  $\zeta_t: \bar{X}_t \times \bar{Y}_t \rightarrow R_m$ , где  $R_m$  – евклидово пространство функционалов. При  $T_{\zeta_t} \equiv T$  контекст не учитывается, т.е. вид трансформации  $T$  не определяется в каждый момент времени предысторией процесса.

При таком определении адаптивного процесса его формирование осуществляется в результате управляемого блуждания в пространстве состояния этого процесса. При этом правилами учитывается  $l$  предыдущих состояний  $\bar{x}_{[l]}$  и управляющих воздействий  $\bar{y}_{[0:l-1]}$ .

Определим модель трансформаций, обеспечивающую целенаправленные переходы в фазовом пространстве. Для этого определим стратегию обработки  $\sigma$ . В этом случае выбор правил управления осуществляется в соответствии с

выбранной стратегией. В этих условиях модель трансформаций приобретает вид:

$$Y_\sigma = \{\sigma, x_o, x_w, DD, D\Phi, DX, DY\}, \quad (2)$$

где  $x_o, x_w$  – это начальное и конечное состояния адаптивного процесса,  $DD \subseteq D$  – множество допустимых правил переходов,  $D\Phi \subseteq \Phi$  – множество допустимых функционалов,  $DX \subseteq X$  – множество допустимых состояний фазового пространства,  $DY \subseteq Y$  – множество допустимых управлений. Все множества определены для выбранной стратегии  $\sigma$ .

Адаптивные свойства процессов могут быть существенно улучшены за счет определения и перестройки допустимых множеств на каждом шаге формирования процесса. Раскроем модель трансформаций для этого случая на момент времени  $k$ :

$$\begin{aligned} Y_{\sigma,k} &= \{x_{k-1}, y_k(x_{k-1}), F_k(x_{k-1}), \\ &DD_k(x_{k-1}), D\Phi_k(x_{k-1}), DX_k(x_{k-1}), DY_k(x_{k-1})\} \\ DD_k &\subseteq DD, D\Phi_k \subseteq D\Phi, DX_k \subseteq DX, DY_k \subseteq DY. \end{aligned} \quad (3)$$

Для перестраиваемых множеств имеют место следующие зависимости. Управляющее воздействие на шаге  $k$  зависит от состояния объекта управления, которое является результатом предыдущих управлений. Выбранная стратегия  $\sigma$  определяет допустимое множество правил управления. В свою очередь, правила управления ограничивают возможные воздействия, оказываемые на объекты управления. Как следствие, сужается допустимое множество состояний в фазовом пространстве, в которые объект может перейти.

Все приведенные модели применимы на практике. Целесообразность их применения определяется свойствами процессов адаптивной обработки, в первую очередь, устойчивостью этих процессов. В случае, когда при изменении условий обработки сами процессы практически не изменяются, может успешно применять модель  $Y$ . Если же процессы изменяются даже при незначительных изменениях контента или контекста, то можно считать обоснованным использование модели  $Y_{\sigma,k}$ .

#### IV. МЕТОД ПОСТРОЕНИЯ ПРОЦЕССОВ ТРАНСФОРМАЦИИ ДАННЫХ

Модели трансформаций (1) – (3) позволяют строить параметрические семейства управляемых процессов трансформации данных, обеспечивающих достижение конечной цели на классе  $K \times \Phi \times \Pi$  за  $m$  итераций. При этом процессы трансформации имеют следующую структуру:

$$F_k = T_{\zeta_k} \otimes F_{k-1}, k = \overline{1, m}, \quad (4)$$

где  $F_k$  – правила, применяемые на  $k$ -ом шаге;  $T_{\varepsilon_k}$  – трансформации, определенные для  $k$ -ого шага;  $\otimes$  – операция трансформации. В соответствии с (4) контент  $\xi$  порождает управляемый процесс обработки. Отметим, что одной из основных целей при построении адаптивных процессов обработки данных является минимизация  $k$ .

Для построения процессов трансформации данных предлагается использовать следующий метод.

1. Определение исходного пространства построения адаптивных процессов. При этом формируются множества:  $X, Y, \Phi, D$ .
2. Выбор стратегии  $\sigma$ .
3. Построение множеств  $DX, DY, D\Phi, DD$  для выбранной стратегии  $\sigma$ .
4. Назначение начального и конечного состояний формируемого процесса адаптивной обработки в фазовом пространстве  $X_o, X_w$ .
5. Назначение шага обработки  $k = 1$ .
6. Построение множеств  $DD_k, D\Phi_k, DX_k, DY_k$ .
7. Оценка альтернативных вариантов управляющих воздействий с использованием функционалов из множества  $D\Phi_k$ .
8. Выбор и реализация управляющего воздействия  $u_k$ , в результате которого управляемый процесс адаптивной обработки переходит в состояние  $k + 1$ , т.е.  $x_k \xrightarrow{u_k} x_{k+1}$ .
9. Оценка соответствия текущего состояния требуемому. Если соответствие установлено, то завершение работы. Иначе, проверяются условия продолжения работы алгоритма. Если условия выполняются, то  $k = k + 1$ , переход на шаг 6.

Условия продолжения работы алгоритма, как правило, основываются на определении максимально допустимого числа шагов  $k_{\text{дон}}$  или максимально допустимого времени обработки  $\tau$ .

Представленный метод имеет достаточно низкую вычислительную сложность относительно других существующих алгоритмов поиска оптимальных путей в фазовом пространстве [5]. При наличии таксономий, систематизирующих элементы допустимых множеств, основная часть операций является операциями поиска элементов по дереву. Сложность таких операций для бинарных деревьев составляет в худшем случае  $O(n)$ . Учитывая, что между числом элементов множеств  $D, \Phi, X, Y$  обычно наблюдается зависимость  $N(D) < N(Y) < N(\Phi) < N(X)$ , то при коэффициенте  $2^k$  общая сложность составляет  $O(n + 2n + 4n + 8n)$ .

## V. ЗАКЛЮЧЕНИЕ

Предложенный метод обеспечивает адаптивную обработку данных, поступающих от технических объектов, за счет построения процессов трансформации этих данных. Метод имеет достаточно низкую вычислительную сложность.

Дальнейшее развитие метода предполагает описание моделей трансформаций в терминах теории категорий [6] и построение автоматных моделей для описания процессов трансформации данных.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Назаров А.В. Современная телеметрия в теории и на практике. Учебный курс / А.В. Назаров, Г.И. Козырев, И.В. Шитов и др. СПб.: Наука и Техника, 2007. 672 с.
- [2] Бусленко Н.П. Моделирование сложных систем. Изд. 2-е. М. Наука 1978 г. 400 с.
- [3] Steinberg A, Bowman C., White F. Revisions to the JDL data fusion model Proc. SPIE Vol. 3719, p. 430-441, Sensor Fusion: Architectures, Algorithms, and Applications III, Belur V. Dasarathy. 1999.
- [4] Срагович В.Г. Теория адаптивных систем. Москва: Наука, 1976. 317 с.
- [5] Zaki M., Meira W., Data Mining and Analysis: Fundamental Concepts and Algorithms, Cambridge University Press, 2014.
- [6] Кнут Д.Э. Искусство программирования. Том 1. Основные алгоритмы = The Art of Computer Programming. Volume 1. Fundamental Algorithms / под ред. С. Г. Тригуб (гл. 1), Ю. Г. Гордиенко (гл. 2) и И. В. Красикова (разд. 2.5 и 2.6). 3. Москва: Вильямс, 2002. Т. 1. 720 с.
- [7] Thorsen S. N., Oxley M. E. Describing data fusion using category theory // Proceedings of the Sixth International Conference on Information Fusion, ISIF, Cairns Australia. 2003. С. 1202-1206.