

В. И. Батищев, Н. Г. Губанов

Методы адаптивного формирования информационных систем анализа состояния сложных технических объектов

Работа посвящена методам автоматической структуризации данных и формирования категорных моделей баз знаний, построению и исследованию методов формирования многомодельных комплексов, основанных на синтезе дедуктивных и индуктивных методов логического вывода.

Особенности проблемы повышения эффективности информационных систем анализа состояния сложных объектов

Создание комплексной методологии анализа и построения информационно-аналитических систем оценки состояния сложных технических объектов является актуальной задачей в русле современных тенденций к интеграции информационно-измерительных систем, систем имитационного моделирования, систем интеллектуального анализа данных, а также подсистем управления базами данных и базами знаний. Данный факт подтверждается активными работами по созданию, внедрению и техническому сопровождению автоматизированных систем четвертого поколения, которые характеризуются как адаптивные интегрированные пространственно-распределенные неоднородные системы обработки данных с перестраиваемыми структурами [1]. Необходимость создания подобных методологий обусловлена всевозрастающей сложностью и стоимостью объектов анализа — сложных технических объектов, когда число оцениваемых параметров исчисляется десятками тысяч, а также накладываются топологические, структурные аспекты при обработке информации, необходимость вычислений в реальном масштабе времени. Кроме того, объекты анализа данного класса характеризуются единичностью изготовления, разнородностью, неполнотой, а зачастую противоречивостью данных об объекте, с одной стороны, и информационной избыточностью с другой.

Указанные свойства обуславливают ряд объективных проблем в вопросах эффективного принятия решений на этапах целевого использования объекта анализа. Специалисты [2] указывают на неточность исходных данных как на основную причину неточности анализа состояния сложных систем. Неполнота и противоречивость данных о системе обусловлена дороговизной, неэффективностью, а зачастую и невозможностью получения полной информации об объекте и среде его функционирования, разнородностью информации об объекте в виде: точечных замеров и значений параметров; допустимых интервалов их изменения; статистических законов распределения для отдельных величин; нечетких критериев и ограничений, полученных от специалистов-экспертов.

В качестве некоторого обобщения можно назвать следующие основные источники, формирующие информационное пространство: данные на выходе ИИС; известные закономерности — заложенные в техдокументации, где данными являются объективные законы реального мира, накопленные в фактографических и документальных системах; выявленные закономерности, в частности имитационные модели. Каждый из источников в настоящее время является информационной основой для соответствующих направлений системного анализа, моделирования и управления сложными системами. Однако каждый вид ресурса обладает рядом принципиальных ограничений, существенно сужающих область его применения. В то же время есть существенные

предпосылки для системной интеграции перечисленных ресурсов. Применение комбинации подходов правдоподобного и достоверного вывода позволит получать новые нелинейные эффекты при синтезе информационно-аналитических систем.

Возникает необходимость в конструктивном формальном аппарате, инвариантном к представлению и обработке разнородной информации из вышеперечисленных источников.

Комплексный подход к формированию алгоритмов логического вывода для анализа состояния сложных объектов

При построении систем анализа, как правило, встают вопросы выбора методов формализации информационных ресурсов, формирования правил построения структур, а также анализ и построение на данных структурах алгоритмов генерации и отбора альтернатив решений. Конструктивным подходом к интеграции различных видов моделей объекта является применение методов категорно-функторного анализа. Данный подход позволяет сохранить целостность представления объекта за счет инвариантности способа полимодельного описания объекта и свести исследование задач одного вида к задачам другого вида, а согласование разнородных моделей осуществлять на основе анализа принадлежности к заданной категории. Предложены алгоритмы автоматического формирования категорных структур, основанные в частности на основе анализа мер близости. В рамках данного подхода разработаны правила формирования полимодельных структур на основе операций наследования и композиции.

Практика показала эффективность синтеза различных подходов вывода в системах анализа. В основу построения баз знаний заложен синтез индуктивных и абдуктивных методов логического вывода. Абдукция как процесс формирования объясняющей гипотезы, служит методологической основой построения алгоритмов правдоподобного вывода. Функционально абдуктивный вывод заключается в принятии решения по выбору опти-

мального объяснения наблюдения на основе заданной теории. Для данного исследования абдукция интересна как средство решения следующих задач: распознавания целей и стратегий деятельности субъекта, формирования моделей по наблюдениям за объектом, накопления и усвоения знаний.

Алгоритм функционирования систем для пересматриваемой аргументации включает следующие процедуры [3]: определение аргумента как дерева выводов, основанного на посылках, либо как дедукции; определение конфликта между аргументами, которые идентифицируются как опровержение аргумента; определение поражения аргумента формированием бинарного отношения на множестве аргументов; оценка аргументов по параметрам, определяемым спецификой предметной области.

Отмечается [4, 6], что по настоящее время количественная оценка гипотез и, соответственно, сравнение и отбор закономерностей являются, во многом, нерешенными проблемами. Проблемы построения абдуктивного вывода заключаются в выборе критерия оценки варианта объяснения, характеризующего степень его правдоподобия. Общая оценка качества абдукции базируется на ряде подходов: подходы, основанные на покрытии множеств, предполагающие порождение и отбор подмножества гипотез, представляющих наилучшее объяснение для наблюдения; подходы, основанные на логике; подход на уровне знаний, рассматривающий абдукцию на уровне неявных убеждений; подходы, основанные на логических моделях.

Индуктивный вывод, в сложных системах позволяющий строить обобщенные модели знаний, основан на построении некоторого общего правила и анализа конечного множества наблюдаемых фактов. Качество обобщенных моделей зависит от полноты набора фактов, которым он пользуется при формировании гипотез. Процедурно процесс индуктивного вывода сложно формализуем и заключается в машинном построении новых гипотез на основе наблюдаемых фактов. Индукционный вывод позволяет решать следующие классы задач: задача индуктивного формирования по-

нятий с целью выделения наиболее общих или характерных фрагментов знания, избавляясь от случайной несистемной информации; задача машинного обучения, где на основе анализа обучающей выборки дается прогноз о новых объектах; задача распознавания, заключающаяся в формировании решающего правила, относящего объект к определенному классу.

В настоящее время существуют конструктивные методы автоматического формирования алгоритмов мониторинга и управления структурной динамикой сложных технических объектов (СТО) [1,6], которые базируются на оперативном формировании операторных цепочек, последовательном отображении измерительных и вычислительных параметров СТО для достижения цели анализа или управления. Данные технологии в частности базируются на обобщенных вычислительных моделях, являющихся развитием недоопределенных моделей.

На принципах индуктивного логического вывода, статистической обработки информации, а также информационных технологиях DM, OLAP, KDD основан целый класс аналитических систем, которые, по мнению аналитиков рынка программных продуктов, составляют существенную часть стоимости СТО в целом. Полученные современные результаты и алгоритмы, позволившие автоматизировать решение таких задач, как: синтез схем программ как доказательство теоремы в формально-дедуктивной системе, методология проверки на модели (model checking), алгоритмы распараллеливания вывода; определения критерия качества моделей в индуктивном выводе на основе теоретико-информационного подхода в русле принципа минимальной длины описания через понятие алгоритмической сложности — показывают перспективность использования в рамках одной системы комплексного подхода, включающего абдукции для получения гипотез, объясняющих наблюдения за параметрами системы состоянием среды, индукции для формирования и оценки правил вывода, и дедукции — для прогнозирования перспективного состояния системы.

Формальная модель информационных систем анализа состояния сложных объектов на основе комплексных алгоритмов логического вывода

Приняв во внимание, что составляющие системы — программные объекты, и взяв за основу систему продукций, рассмотрим следующую формальную модель:

$$P_i = \langle M, M', R_i, O_i \rangle,$$

где в качестве множества заданных литералов M_i продукционной системы и множества формируемых литералов M'_i продукционной системы определены обобщенные вычислительные модели;

R_i — множество продукций i -го вида;

O_i — множество процедур присвоения i -го вида.

$$M = \langle A, F_M \rangle,$$

где $A = \{\alpha_i, i = 1, \dots, n\}$ — конечное множество параметров состояния объекта;

F_M — конечное множество отношений на множестве параметров из A ;

$F = \langle f_i, i = 1, \dots, k \rangle | a \in A$ — отношение на множестве параметров $f = \langle A, gr(f) \rangle$.

Множество всех отображений ω для всех отношений $f \in F_M$, входные $in(\omega) = Z_A^{in}$ параметры для оператора ω , выходные $out(\omega) = Z_A^{out}$ параметры для оператора ω . Взяв за основу категорный подход к формированию продукций [5], можно выстроить иерархию моделей заданной категории по степени детализации: M_i^B — базис моделей i -категории, (M_i^0, ψ) — исходное состояние модели (вычислительного алгоритма), $\psi \in Hom(S_i^0, S_i^B)$, — производное исходное состояние модели (вычислительного алгоритма), $\lambda \in Hom(M_i^u, M_i^B)$ — условие сопоставимости. Соответственно, распознавая ситуацию, система активизирует некоторую продукцию, сопоставимую с заданной ситуацией. Специалисты указывают [6], что в данной ситуации актуальна задача квалиметрии моделей, которая заключается в формировании подхода, позволяющего с единых позиций проводить оценку, сравнение, упорядочивание моделей. Сформированное

множество иерархий моделей позволяет использовать их в алгоритмах вывода. Отбор моделей осуществляется на основе двух альтернатив: при наличии удовлетворительной модели детализировать выходные данные (дедуктивные алгоритмы вывода); при отсутствии удовлетворительной модели построить обобщенную модель (индуктивные алгоритмы вывода).

Как было упомянуто ранее, особый интерес представляет не просто оценка параметров системы, а анализ топологии, соответствующей структуре объекта. В частности, конструктивна идея доказательства непрерывности отображения между физическими процессами в системе, измерительной информацией и состоянием вычислительного процесса, что позволяет говорить об адекватности процесса анализа.

Основу систем анализа состояния СТО составляют полимодельные комплексы. Проблема создания алгоритмов формирования и представления полимодельных структур является одной из ключевых в современном системном моделировании. Структура и функционирование аналитических систем зависят от следующих информационных сущностей: объекта анализа (СТО) — Q ; цели функционирования аналитической системы — G , определяемой конкретной задачей принятия решения; полимодельного комплекса, задающего структуру системы — M ; среды, определяющей параметры системы — C , а также отношений между данными структурами $R = (r_{Q,M}, r_{Q,C}, r_{Q,G}, r_{M,C}, r_{Q,M}, r_{Q,C})$. Соответственно, информация по всем имеющимся в распоряжении субъекта информационным ресурсам Z определяется как $Z = \bigcup_Q (M, R)$, тогда формирование новой структуры информационно-аналитической системы представим в следующем виде:

$$F = (Z, M_G, \theta),$$

где M_G — целевая структура системы;

$\theta = (M_b, p)$ — алгебра формирования структур,

где M_b — множество базовых классов элементов структур M ;

$p = (N, K)$ — операции формирования структуры системы,

где N — операция наследования;

K — операция композиции.

Комбинация данных операций формирования структуры системы, в отличие от конкатенации, позволяет сохранять целостность представления системы на различных уровнях иерархии. Алгоритмы формирования базовых классов объектов $A_b = (Z, G, M_b)$ являются, по сути, проблемно-ориентированной декомпозицией Z , стратегия построения данных алгоритмов лежит в русле принципа семиотической интроспекции, заключающегося в идентификации различий и обобщении подобий множества объектов.

Модель объекта M^i описывает его некоторые свойства в соответствующих категориях. Соответственно, можно рассматривать M^i как объект категории ObM^i , а взаимосвязь между объектами — как морфизмы $MorM^i$. Применительно к задаче таксономии, категории формируются на основании [7] признакового пространства $I^n = \{i_1, i_2, \dots, i_n\}$, множества классов $KI(M)$ и самих объектов таксономии M . Объекты, принадлежащие одному классу, являются изоморфными, другими словами, неразличимыми в признаковом пространстве I^n , а классы объектов $KI(M)$ в данном признаковом пространстве являются гомоморфными, образуя при соответствующих свойствах признакового пространства категорию $Cat(M)$.

Полимодельное описание объекта определяется совокупностью моделей различных категорий $M^{poly} = \bigcup_{i=1}^k M^i$. Формирование правил отображения модели одного вида в другой требует построения функтора $F(M^v, M^u)$, вид которого определяет вид отношений между моделями видов $\langle v, u \rangle$. Соответственно, возможно формирование знаний категории $Cat(M_u)$ при недостаточных условиях формирования категории с помощью процедуры таксономии A возможна процедура:

$$A': \exists F(M^v) \rightarrow Cat(M^u),$$

которая позволяет строить и обрабатывать гипотезы относительно знаний одной катего-

рии, применяя их к знаниям другой категории, что позволяет расширить практические возможности добывания знаний.

Алгоритмы реструктуризации данных полимодельных комплексов, заключающиеся в коррекции отношений между объектами категории, формировании новых категорий и редакции имеющихся, основаны на подходе наследования новой структуры из существующей, формирования иерархической или сетевой структуры из компонент полимодельных комплексов.

Пусть S_b^{pm} — базис категории, (S_0^{pm}, ψ) — исходный объект, отношение между данными объектами — $\psi \in Hom(S_0^{pm}, S_b^{pm})$, (S_i^{pm}, λ) — производный объект, где $\lambda \in Hom(S_i^{pm}, S_b^{pm})$.

Условие сопоставимости следующее:

$$\exists \omega \in Hom(S_0^{pm}, S_i^{pm});$$

$$\exists (\mu \in Hom(S_i^{pm}, S_j^{pm})) \rightarrow ((S_i^{pm}, \varphi)N(S_j^{pm}, \psi)),$$

тогда

$$\exists (\mu \in Hom(S_i^{pm}, S_j^{pm})) \rightarrow ((S_i^{pm}, \varphi)N(S_j^{pm}, \psi)),$$

где N — операция наследования объекта (S_i^{pm}, φ) в (S_j^{pm}, ψ) .

В общем виде стратегия оперирования вычислительными моделями следующая: либо модель является объектом дедуктивного вывода и формирует более детальный результат, либо она является объектом индуктивного для структур более высокой иерархии.

Таким образом, задание цели анализа вызывает активизацию нескольких конкурирующих вычислительных схем $Y = (\Gamma_1, \Gamma_2, \dots, \Gamma_n)$, формирование которых осуществляется на основе $I(Y, \gamma_j^k)$ множества игровых ситуаций.

Заключение

Вышеизложенный материал — основа формально-математического аппарата для представления, анализа и обработки разнородных данных и моделей, включающий в себя методы автоматической структуризации данных и формирования категорных моделей баз знаний, методы анализа свойств и отношений на

моделях данных и знаний, методы формальной декомпозиции и агрегирования на элементах структур данных. Данная методология показала свою эффективность в процессах автоматизации формирования модели городской транспортной системы.

СПИСОК ЛИТЕРАТУРЫ

1. Охтилев М. Ю., Соколов Б. В., Юсупов Р. М. Интеллектуальные технологии мониторинга и управления структурной динамикой сложных технических объектов. М.: Наука, 2006.
2. Батищев В. И., Мелентьев В. С. Аппроксимационные методы и системы промышленных измерений, контроля, испытаний, диагностики. М.: Машиностроение-1, 2007.
3. Достоверный и правдоподобный вывод в интеллектуальных системах / В. Н. Вагин, Е. Ю. Головина, А. А. Загорянская [и др.]. М.: Физматлит, 2004.
4. Батищев В. И., Губанов Н. Г. Категорное представление сложных технических объектов в индуктивных системах логического вывода // Проблемы управления и моделирования в сложных системах: труды IX Международной конференции. Самара: СНЦ РАН, 2008.
5. Стефанюк В. Л. Локальная организация интеллектуальных систем. М.: Физматлит, 2004.
6. Соколов Б. В., Юсупов Р. М. Концептуальные основы оценивания и анализа качества моделей и полимодельных комплексов // Известия РАН. Теория и системы управления. 2004. № 6. С. 5–16.
7. Загоруйко Н. Г. Прикладные методы анализа данных и знаний. Новосибирск: Изд-во Ин-та математики.
8. Потапов А. С. Распознавание образов и машинное восприятие. Общий подход на основе принципа минимальной длины описания. СПб.: Политехника, 2007.
9. Губанов Н. Г. Категорный подход при формировании полимодельных комплексов сложных систем // Вестник Самар. гос. техн. ун-та. 2008. Вып. 1(21). С. 183–185. (Технические науки).
10. Батищев В. И., Губанов Н. Г. Методология оперативной реструктуризации информационных систем анализа состояния сложных технических объектов [Текст] // Проблемы управления и моделирования в сложных системах: труды IX Международной конференции Самара: СНЦ РАН, 2008. С. 187–193.