

оброблення таких сигналів, відбуваються процеси зворотного перетворення. Детермінованість частотно-часових послідовностей, що сформовані за допомогою розробленого способу, дозволяє використовувати під час прийому таких сигналів кореляційні методи обробки і тим самим робить можливим усунення захисного інтервалу між символами повідомлення.

Висновки. Розроблена технологія побудови передавальної й приймальної частини систем бездротового доступу, фізичний рівень яких сформований на основі застосування багаточастотних кодових конструкцій без збігу частотно-часових елементів, дозволяє за рахунок використання на фізичному рівні частотно-часового кодування збільшити швидкість передачі інформації при заданій достовірності, тим самим підвищити інформаційну ефективність систем бездротового доступу.

Виведені нові аналітичні вирази дозволяють без значних складностей на практиці програмним або апаратно-програмним методами реалізувати розроблену технологію.

Оцінювання, проведене в роботі [3], підтверджує доцільність впровадження розробленої технології в системах бездротового доступу, які на фізичному рівні використовують складні багаточастотні сигнальні конструкції.

Список використаних джерел

1. Яриловець А. В. Аналіз стану та перспективи розвитку телекомунікаційних мереж / А. В. Яриловець, С. В. Зайцев, В. Д. Назарук // Вісник Черніг. держ. технол. ун-ту. – 2012. – № 2 (57). – С. 155–165.
2. Яриловець А. В. Алгоритм побудови оптимальних частотно-часових сигнальних конструкцій / А. В. Яриловець, В. Д. Назарук, С. В. Зайцев // Математичні машини і системи. – 2012. – № 4. – С. 94–102.
3. Яриловець А. В. Оцінка ефективності інформаційної технології для систем безпроводового доступу, побудованої на основі оптимальних ортогональних багаточастотних сигнальних конструкцій / А. В. Яриловець, В. Д. Назарук, С. В. Зайцев // Математичні машини і системи. – 2014. – № 2. – С. 51–58.

УДК 650.1

А.А. Задорожний, асистент

Черниговский национальный технологический университет, г. Чернигов, Украина

КЛАССИФИКАЦИЯ И МЕТОДЫ УСТРАНЕНИЯ НЕОПРЕДЕЛЕННОСТЕЙ В МОДЕЛЯХ ПРИ ИНДУКТИВНОМ И ДЕДУКТИВНОМ ПОДХОДАХ К ИХ СОЗДАНИЮ

А.О. Задорожній, асистент

Чернігівський національний технологічний університет, м. Чернігів, Україна

КЛАСИФІКАЦІЯ І МЕТОДИ УСУНЕННЯ НЕВИЗНАЧЕНОСТЕЙ У МОДЕЛЯХ ПРИ ІНДУКТИВНОМУ І ДЕДУКТИВНОМУ ПІДХОДАХ ДО ЇХ СТВОРЕННЯ

Artem Zadorozhniy, assistant

Chernigov National Technological University, Chernigov, Ukraine

CLASSIFICATION AND METHODS FOR RESOLUTIONS OF INDETERMINACIES IN THE MODELS CREATED WITH INDUCTIVE AND DEDUCTIVE APPROACHES

Проведен анализ неопределенностей, которые могут возникать при создании моделей процессов и систем. Рассмотрены существующие классификации неопределенностей и предложена расширенная классификация, которая включает не только неопределенности, связанные данными и внутренними состояниями модели, но и неопределенности, связанные с жизненным циклом построения модели.

Ключевые слова: неопределенность, индуктивный подход, дедуктивный подход, жизненный цикл модели.

Проведено аналіз невизначеностей, що можуть виникати при створенні моделей процесів і систем. Розглянуто наявні класифікації невизначеностей та запропоновано розширену класифікацію, що включає не тільки невизначено-

сті, пов'язані з даними та внутрішніми станами моделі, але й невизначеності, пов'язані з життєвим циклом побудови моделі.

Ключові слова: невизначеність, індуктивний підхід, дедуктивний підхід, життєвий цикл моделі.

The analysis of the indeterminacies that may occur during creating moles of processes and systems has been made. The existing classifications of indeterminacies have been researched and expanded classification, which includes not only the indeterminacies associated with data and internal states of the model, but also the indeterminacies associated with the life cycle of model creation, has been proposed.

Key words: indeterminacy, inductive approach, deductive approach, model life cycle.

Постановка проблеми. Во время создания моделей систем и процессов могут возникать различного рода неопределенности, которые необходимо устранять. Существующие классификации неопределенности, связанные с входными данными модели и ее внутренними состояниями, в то время как неопределенности, связанные с жизненным циклом создания модели, также являются важным элементом в классификации неопределенностей. Также в данной работе рассматриваются способы устранения неопределенности внутренних состояний моделей для индуктивного и дедуктивного подхода к созданию имитационных моделей.

Анализ последних исследований и публикаций. Изучение проблем неопределенностей в имитационном моделировании приобрело широкое распространение в современной науке. Разрабатываются классификации неопределенностей [1], а также предлагаются способы их решения. Наиболее хорошо изучены и обеспечены методами решения неопределенности, связанные с входными данными и внутренними состояниями модели.

Выделение не решенных раньше частей общей проблемы. Нерешенной частью проблемы классификации неопределенностей является отсутствие в существующих классификациях неопределенностей, связанных с жизненным циклом создания моделей. Также одной из проблем в предметной области неопределенностей является отсутствие метода, позволяющего исследовать внутренние состояния индуктивных моделей.

Цель работы. Главная цель данной работы – расширение классификации неопределенностей включением неопределенностей, связанных с жизненным циклом модели, в существующую классификацию, а также рассмотрение методов для устранения неопределенностей внутренних состояний моделей при индуктивном и дедуктивном подходах к созданию моделей.

Понятие и классификация неопределенностей. Неопределенность – это отсутствие или недостаток определения или информации о чем-либо. Неопределенность может проявляться в разных областях, таких как неопределенность измерений в метрологии, принцип неопределенности Гейзенберга в квантовой физике, раскрытие неопределенности в математическом анализе, неопределенность в управлении рисками.

В аспекте проблемы неопределенности принято различать ряд терминологических групп понятий, характеризующих категорию «неопределенность» с разных точек зрения. При этом каждое понятие о неопределенности является связанным с другими и не противоречит содержательному наполнению положений, а только дополняет и расширяет сложившиеся представления.

Неопределенность можно рассматривать как меру информации [1]. Это является первым и самым распространенным положением о неопределенности. Достаточность информации об условиях, ограничениях и параметрах сложных систем и процессов свидетельствует об определенности ситуации. В таком контексте восприятия категории подразумевается, что совокупная информация о конкретном объекте, событии или явлении априори составляет константу полных истинных сведений и данных.

Вторым положением о неопределенности является то, что она отражает состояние системы по отношению к идеальным условиям, когда знание полностью детерминировано. Положение о разрыве между фактическим уровнем «информированности» и си-

туацией, когда сведения и данные о сложной системе или процессе полностью известны очень близко к первому положению о неопределенности и, по сути, является его следствием. Основываясь на первых положениях о неопределенности, объем информации может быть исчислен и выражен через неопределенность – энтропию. Таким образом, различие между физическим содержанием информационных потоков и идеальным объемом истинных сведений и данных характеризует второе положение о неопределенности.

Третьим положением неопределенности является то, что она воспринимается как возможность выбора альтернатив и множественность данного выбора (вариантов выбора). Во многих научных публикациях делается акцент на то, что определенность порождает множественность выбора различных альтернатив. С одной стороны, множественность вызвана разнообразием вариантов, с другой стороны, в условиях неопределенности установить четкие критерии оптимальности и эффективности достаточно сложно.

В четвертом положении неопределенность определяет качество информации (достоверность, полнота, ценность, актуальность, ясность). Анализ понятий неопределенности данного положения показал, что содержательно группа понятий включает большее количество определений, которые тем или иным способом дают оценку информации. Чаще всего оценка информации в контексте неопределенности связана с достоверностью сведений и данных, их полнотой и объективностью.

Пятое положение о неопределенности говорит, что неопределенность является атрибутивным источником риска. Априори риск действительно находится в прямой зависимости от неопределенности, то есть при росте неопределенности возрастает и риск. Величина прироста может меняться и обусловлена эластичностью рисков по отношению к неопределенности. Вопросу взаимосвязи неопределенности и рисков посвящены многие научные работы, в которых утверждается, что неопределенность является прямым источником рисков.

Шестое положение о неопределенности говорит, что неопределенность предполагает неоднозначность реализации событий, порождаемая факторами неизвестной природы. Использование синтетического подхода к исследованию данного положения о неопределенности приводит к тому, что оно очень схоже с третьим положением – возможности выбора альтернатив и множественности данного выбора. Однако суть положения о неоднозначности реализации событий подразумевает результат возникновения каждого события. Решение принимается в условиях неопределенности, когда невозможно оценить вероятность потенциальных результатов. Это должно иметь место, когда требующие учета факторы настолько новы и сложны, что насчет них невозможно получить достаточно релевантной информации. Таким образом, неопределенность и создает множественность результатов, которые в дальнейшем подвергаются взвешенной оценке при анализе рисков с использованием математического ожидания и других средств усреднения.

В седьмом положении неопределенность является естественным ограничением управляемости и стабильности сложной системы или процесса. Энтропия является параметром оценки неопределенности в сложных системах и процессах. В общем случае, энтропию можно охарактеризовать как меру хаоса, беспорядка. Обратным проявлением энтропии является негэнтропия и ее суть обратно пропорциональна энтропии. Она характеризует упорядоченность системы. Отношение энтропии и негэнтропии описывает возможность системы быть в устойчивом, стабильном состоянии. Это состояние говорит об управляемости системы. Чем выше управляемости системы, тем система является более гибкой к воздействиям внешней и внутренней среды. Таким образом, в сложных системах и процессах существует порог управляемости и стабильности, который

детерминирован неопределенностью. С этим связано возникновение сопряженных теорем: теоремы предельной управляемости (неопределенности) и теоремы предельной самоорганизации (стабильности).

Большинство сложных систем и процессов состоят из совокупности целого ряда взаимосвязанных многорежимных подсистем, функционирование которых направлено на достижение общих целей. В свою очередь для каждой подсистемы может быть определена цель ее функционирования. Между элементами системы могут существовать внутренние материальные, энергетические и информационные связи, а также система может иметь связи с другими объектами и в ее работе могут участвовать люди и машины и внешняя среда.

Большинство систем, в которых протекают различные процессы, функционируют в условиях неопределенности. Эта неопределенность возникает по ряду причин, основными из которых являются:

- сложность объекта и недостаточная изученность процессов, протекающих в нем;
- стохастическая природа основных параметров, описывающих функционирование объекта либо процесса;
- наличие большого количества возмущающих воздействий и помех, зашумленность информации;
- недостаточная достоверность исходной статистической информации, обусловленная низкой надежностью, нехваткой или отсутствием средств сбора и обработки информации;
- недостатки методов обработки информации.

Один из вариантов классификации неопределенности представлен на рис. 1.

Таким образом, основными источниками неопределенности в моделях сложных систем являются недостаточность и нечеткость исходной информации. От правильности и удачности выбора модели во многом зависит весь ход дальнейшего исследования сложной системы или процесса.



Рис. 1. Классификация неопределенностей

Конечно, при построении модели желательно получить ее четкое описание. Однако попытки получить четкое описание сложной системы или процесса в тех случаях, когда иная информация о них, кроме нечеткой, недоступна, путем задания строгих границ искусственным введением однозначности, приводит к огрублению исходных данных, которое может способствовать получению четких, но не адекватных моделей и поэтому это не целесообразно. Тем не менее и в этом случае целесообразно попытаться выде-

лить в нечеткой в целом системе или процессе отдельные ее части, поддающиеся четкому описанию.

Таким образом, уже в самом начале процесса построения модели следует ответить на вопрос, какой тип модели четкий или нечеткий может быть получен, исходя из имеющейся информации о системе или процессе в целом или об их составляющих частях.

Неопределенности внутренних состояний модели. Одним из видов неопределенностей, который относится к неопределенностям, связанных со сложностью объекта и недостаточной изученностью процессов, протекающих в нем, является неопределенность внутренних состояний объекта. Такая неопределенность затрудняет применение дедуктивного подхода при создании объекта, поскольку такой подход требует досконального знания процессов, протекающих в объекте наблюдения. Решением при создании моделей систем и процессов, внутренние состояния которой являются неопределенными, является применение индуктивного подхода. Суть индуктивного подхода заключается в использовании данных наблюдений за объектом или процессом для построения их математической модели без учета внутренних состояний объекта. Если говорить более точно, то внутренние состояния на самом деле учитываются, но они скрыты внутри самой математической модели, и исследователь не может делать никаких предположений о состояниях системы или процесса на основании их математических моделей. Для того, чтобы исследовать внутренние состояния сложной системы или процесса, необходимо воспользоваться предложенным в [2] модифицированным алгоритмом метода группового учета аргументов.

Данный метод отличается от стандартного метода группового учета аргументов тем, что с его помощью можно строить не функциональные (математические) модели, базовыми строительными элементами которых являются полиномы, а блочные модели, базовыми элементами которых являются автоматы. Блочные модели могут быть представлены детерминированными либо недетерминированными автоматами-распознавателями и преобразователями. Самым главным в применении этого модифицированного метода является то, результатом является автомат, а значит можно определить количество состояний в модели, а также можно исследовать, какую роль играет данное состояние в том или ином сложном процессе либо системе.

Ниже более детально рассмотрен модифицированный алгоритм метода группового учета аргументов.

Блочные модели могут быть представлены как автоматами-распознавателями, так и автоматами-преобразователями, но наиболее удобным для моделирования является автомат-преобразователь, поскольку у него есть множество как входных, так и выходных состояний, а у автомата-распознавателя выходным является множество принимающих состояний.

Конечный детерминированный автомат-преобразователь – это система вида:

$$A = \langle \Sigma_x, \Sigma_y, Q, q_0, \Phi, \Psi \rangle, \quad (1)$$

где $\Sigma_x = \{a_1, \dots, a_m\} (m \geq 1)$ – конечное множество – входной алфавит;

$\Sigma_y = \{b_1, \dots, b_r\} (r \geq 1)$ – конечное множество – выходной алфавит;

$Q = \{q_0, \dots, q_{n-1}\} (n \geq 1)$ – конечное множество – алфавит внутренних состояний;

$q_0 \in Q$ – начальное состояние автомата;

$\Phi: Q \times \Sigma_x \rightarrow Q$ – функция переходов, $\Phi(q, a)$ – это состояние, в которое переходит автомат из состояния q , когда получает на вход символ a ;

$\Psi: Q \times \Sigma_x \rightarrow \Sigma_y$ – функция выходов, $\Psi(q, a)$ – это символ из Σ_y , который выдает на выход автомат в состоянии q , когда получает на вход символ a .

В методе группового учета аргументов (МГУА) математическая модель процесса представлена в виде многочлена, который получают из исходного простого многочлена добавлением новых членов либо возведением уже существующих членов в степень. Необходимо найти эквиваленты таких операций для структурных моделей.

В нашем случае структурные модели представлены в виде автоматов. И, таким образом, эквивалентами операций суммы и произведения могут являться теоретико-множественные и алгебраические операции, применимые для автоматов.

В качестве эквивалентной операции для усложнения модели можно, например, использовать операцию объединения нескольких автоматов в один более сложный автомат.

Пусть у нас есть автоматы A и B . Предположим, что $A = (X_1, Q_1, Y_1, q_1 \in Q_1, F_1(x \in X_1 / y \in Y_1))$, а $B = (X_2, Q_2, Y_2, q_1 \in Q_2, F_2(x \in X_2 / y \in Y_2))$, тогда автомат $C = (X, Q, Y, q_1 \in Q, F(x \in X / y \in Y))$ является объединением автоматов A и B , если множества X , Q и Y и отображение F определяются по формулам:

$$X = (\{1\} \times X_1) \cup (\{2\} \times X_2), \quad (2)$$

$$Q = Q_1 \cup Q_2, \quad (3)$$

$$Y = (\{1\} \times Y_1) \cup (\{2\} \times Y_2), \quad (4)$$

$$Fq = F_1q \cup F_2q, \quad (5)$$

где $q \in Q$. В том случае, когда $q \notin Q_1$, полагаем $F_1q = \emptyset$, а при $q \notin Q_2$ имеем $F_2q = \emptyset$.

Если

$$X_1 \cap X_2 = \emptyset \quad (6)$$

и

$$Y_1 \cap Y_2 = \emptyset, \quad (7)$$

то сумму множеств в (5) и (7) можно заменить объединением множеств:

$$X = X_1 \cup X_2, \quad (8)$$

$$Y = Y_1 \cup Y_2. \quad (9)$$

Пример усложнения модели путем объединения двух автоматов более детально рассмотрен в [2]. Кроме операции объединения двух автоматов, существует еще операция пересечения двух автоматов, но эта операция не подходит для усложнения структурной автоматной модели, поскольку она упрощает структуру модели, а не усложняет ее. Над автоматами можно еще производить алгебраические операции, такие как умножение, суммирование, суперпозиция и композиция [3], что также можно использовать для усложнения блочных моделей. Модифицированный алгоритм метода группового учета аргументов направлен на определение внутренних состояний блочной модели для последующего их изучения. В дедуктивном подходе существует иная ситуация, когда необходимо знать внутренние состояния модели в процессе ее создания.

Например, при создании моделей с использованием агентного формализма, как и при создании моделей с использованием других формализмов, часто могут возникать неопределенность внутренних состояний модели, что чаще всего связано со сложностями изучения реального объекта. Часто на начальных этапах известны далеко не все данные о состояниях объекта или процесса.

Агентные модели являются представителями объектно-ориентированных моделей, в которых состояния представлены в виде атрибутов объектов. В процессе усложнения модели число таких атрибутов может увеличиваться, что приводит к необходимости изменения программного кода модели. Однако внесение изменений в программный код требует дополнительных накладных расходов на тестирование этих изменений, а также

времени на перекомпиляцию модели. Современные объектно-ориентированные языки, такие как Java и C#, предоставляют механизм рефлексии, который позволяет создавать классы во время выполнения программы без перекомпиляции, что существенно ускоряет процесс добавления новых атрибутов в объекты.

Неопределенность точки завершения усложнения модели. Еще одним видом неопределенностей, который не столько связан с данными, сколько с самим процессом создания модели, является неопределенность точки завершения создания модели.

Неопределенность точки остановки создания модели связана с тем, что процесс создания модели является больше творческим процессом, чем заданной последовательностью действий, по завершению которой будет создана модель. Таким образом, четких этапов создания модели не существует. Кроме того, главными в моделировании являются не сами модели, а эксперименты над этими моделями и выводы по экспериментам. Но количество моделей, которые необходимо создать, со временем увеличивается, и появляются некоторые общие черты при их создании. Особенно хорошо это заметно при создании программных моделей, поскольку жизненный цикл модели при этом схож с жизненным циклом программы, что означает, что на схожих этапах можно использовать одни и те же инструментальные средства. Наиболее удачная последовательность действий была приведена в [4].

Основными этапами создания модели являются формулирование проблемы и содержательная постановка задачи, разработка концептуальной модели, разработка программной реализации модели, которая в свою очередь состоит из выбора средств моделирования, с помощью которых будет реализована модель, разработки структурной схемы модели и составления схемы ее функционирования, программной реализации модели [4]. Завершающими этапами создания модели являются проверка адекватности модели, организация и планирование проведения эксперимента, интерпретация результатов моделирования и принятие решений, оформление результатов моделирования.

Процесс создания модели не является однопроходным, а является итеративным процессом. Каждая итерация моделирования требует от разработчика принять решение, является ли модель окончательной и в достаточной степени схожей с реальным объектом либо процессом, либо необходимо произвести еще одну итерацию с усложнением внутренней структуры модели. При этом необходимы четкие формальные критерии, с помощью которых можно было бы принять решение о необходимости следующей итерации. Такие критерии существуют в индуктивном подходе к моделированию, но нет четких критериев остановки процесса моделирования в дедуктивном подходе. После каждого этапа итерации построения модели необходимо проверить корректность создания концептуальной модели – валидацию и проверить корректность реализации модели – осуществить верификацию модели. Во время проверки достоверности необходимо ответить на вопрос о соответствии модели моделируемой системе, то есть определить, в какой мере модель и система изоморфны. Как правило, в случае моделирования требование изоморфизма модели и объекта чрезмерны, потому что в этом случае сложность модели должна соответствовать сложности объекта. Поэтому строят гомоморфные модели, в которых выполняется условие однозначного соответствия модели объекту. На этапе верификации рассматривают, правильно ли преобразована концептуальная модель (модельные предположения) на компьютерную программу, то есть производят настройку программы моделирования. Это сложное задание, поскольку может существовать множество логических путей. Этап проверки корректности реализации модели включает проверку эквивалентности преобразования модели на каждом из этапов ее реализации и сравнение состояний. В этом случае модель претерпевает таких изменений: концептуальная модель – математическая модель – алгоритм моделирования – программная реализация модели.

Валидация – это процесс, который дает возможность определить, является ли модель (а не компьютерная программа) точным отображением системы для конкретных целей исследований. Разрабатывают план проведения экспериментов с моделью для достижения поставленных целей. Основная цель планирования экспериментов – изучение поведения моделируемой системы при наименьших затратах во время экспериментирования. Наиболее часто проводят следующие эксперименты:

- сравнивают средние значения и дисперсии разных альтернатив;
- определяют важность учета влияния переменных и ограничений, которые накладываются на эти переменные;
- определяются оптимальные значения с определенного множества возможных значений переменных.

Проведение экспериментов планируют для поиска незначимых факторов. В случае оптимизации какого-нибудь числового критерия формируют гипотезу относительно выбора наилучших вариантов структур моделируемой системы или режимов ее функционирования, определяют диапазон значений параметров (режимов функционирования) модели, в пределах которого находится оптимальное решение. Определяют количество реализаций и время прогона модели каждой реализации. Проводят экстремальный эксперимент, по результатам которого находят оптимальное значение критерия и соответствующие значения параметров. Для оценки точности стохастических моделей строят доверительные интервалы для получения выходных переменных.

Проверка модели на адекватность является одним из критериев завершения итерации жизненного цикла модели наряду с модульным и интеграционным тестированием программной реализации модели. Проверка модели на адекватность выполняется в том случае, если было проведено экспериментирование с реальным объектом либо процессом. В случае, если данные наблюдений над реальным объектом либо процессом отсутствуют, необходимо проводить верификацию модели. Проверка модели на адекватность (верификация в случае отсутствия данных наблюдения) похожа по своей природе на тестирование. Задачей тестирования является сравнение выходных данных с эталоном, задачей проверки модели на адекватность является проверка адекватности отклика при заданных входных значениях фактора. В случае верификации модели, когда отсутствуют данные наблюдения за реальным объектом либо процессом, сравнение данных не происходит, а проводится исследование набора откликов модели с использованием набора специальных методов. Таким образом, отличием процесса тестирования модели от процесса проверки ее на адекватность является то, что при проверке модели на адекватность производится проверка наборов входных и выходных данных, а не значений единичного эксперимента. При тестировании чаще всего происходит сравнение выходного результата функции с эталоном. Таким образом, для того чтобы обеспечить автоматизацию процесса проверки модели на адекватность (верификацию), необходимо расширить инструментально средство тестирования дополнительным функционалом для подачи на вход теста наборов данных, например с базы данных, а также функционалом для проведения корреляционного анализа, на основании которого будет применяться решение о адекватности модели.

Неопределенности входных данных моделей. При создании моделей важную роль играют входные данные. При дедуктивном подходе роль этих данных выполняют результаты исследования сложного объекта либо системы, произведенные на основании этих данных выводы, либо личный опыт эксперта, занимающегося созданием модели. При индуктивном подходе входными данным являются выборки наблюдения за системами либо процессами. Сложные системы либо процессы могут вести себя произвольным образом, и их поведение может изменяться во времени по определенным законам.

Время наблюдения за объектом либо процессом может быть ограничено во времени из-за внешних факторов, таких как малое время жизни объекта либо протекания процесса. Это может приводить к появлению неполноты входных данных для создания моделей.

Существует множество способов устранения неполноты данных уже после этапа сбора данных: заполнение средним значением, пропорциональное размещение наблюдений с пропущенными данными по уже имеющимся градациям шкалы, расчет возможного значения при помощи регрессионной модели и так далее. Заполнение пропусков позволяет не только получить дополнительную информацию (предсказание значения), но и сохранить уже имеющуюся, часто очень важную и полученную ценой значительных усилий информацию за счет сохранения наблюдений изначально содержащих пропуски.

Заполнение пропусков данных, кроме достоинств, имеет и несколько недостатков, которые нельзя не учитывать:

- использование для предсказания пропусков имеющихся полных данных искажает структуру результирующих данных, которая смещается в сторону структуры только полных данных;
- искусственная подстановка пропусков вносит в массив определенную долю искусственных данных, которые в свою очередь приводят к смешиванию значимости получаемых на их основании результатов.

Существует большая вероятность того, что модели, построенные по данным с заполненными пропусками, будут менее точными по сравнению с моделью, построенной только на полных данных. Потери в их точности будут зависеть от качества предсказания существующих значений. Но, потеряв в точности, можно выиграть в репрезентативности результатов.

При выборе метода восстановления пропущенных данных следует учитывать, что, так как алгоритмы заполнения пропусков не универсальны, возможность применения того или иного способа заполнения зависит от метода анализа данных, который планируется использовать в дальнейшем.

Наиболее популярные методы устранения неопределенности входных данных отражены на рис. 2.

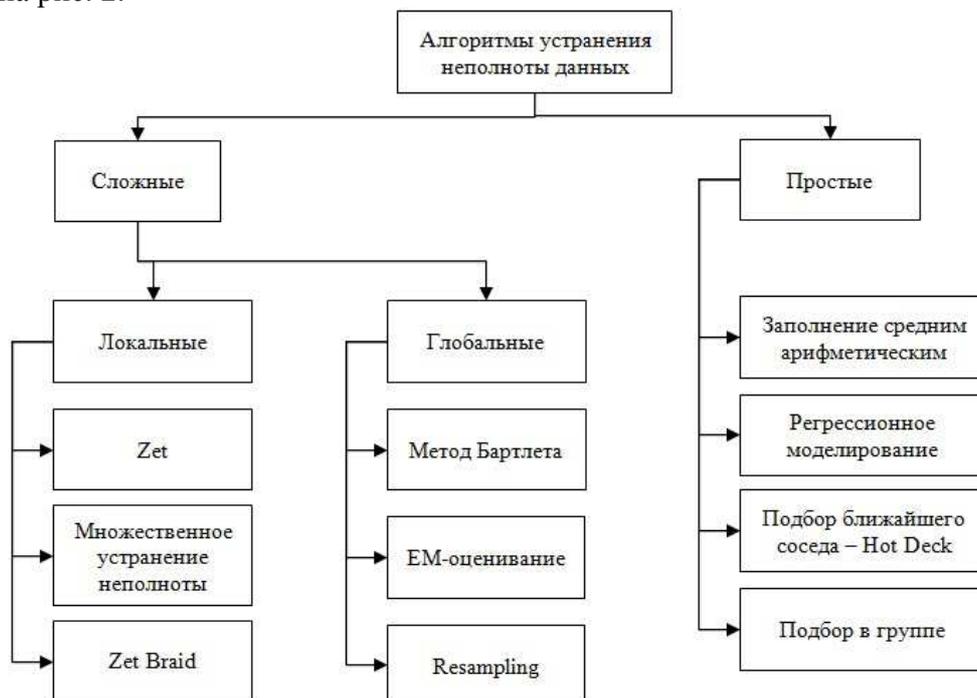


Рис. 2. Классификация алгоритмов устранения неполноты данных

Простые алгоритмы – неитеративные алгоритмы, основанные на простых арифметических операциях, расстояниях между объектами, регрессионном моделировании. К ним относится заполнение пропусков средним арифметическим, регрессионное моделирование пропусков, метод HotDeck и подбор в группе.

Сложные алгоритмы – итеративные алгоритмы, предполагающие оптимизацию некоторого функционала, отражающего точность расчета подставляемых на место пропуска значений. Их можно разделить на глобальные и локальные алгоритмы. В глобальных алгоритмах в оценивании каждого пропущенного значения учувствуют все объекты рассматриваемой совокупности: метод Бартлета, EM – оценивание и Resampling.

В локальных алгоритмах в оценивании каждого пропущенного значения участвуют полные наблюдения, находящиеся в некоторой окрестности предсказываемого объекта. К данной группе относятся алгоритмы Zet и Zet Braid.

Выводы. Процесс создания моделей сложных объектов и процессов часто содержит различного рода неопределенности. Можно выделить такие виды неопределенностей, как неопределенности внутренних состояний модели, неопределенности входных моделей и даже такой вид неопределенности, как неопределенность точки завершения создания модели. Неопределенности внутренних состояний модели можно решить через модифицированный алгоритм метода группового учета аргументов, в котором вместо математической модели производится создание блочной модели на основании автоматных моделей. На основании блочной модели на выходе можно примерно судить о внутренних состояниях модели, а в итоге и о внутренних состояниях сложной системы либо процесса. Неопределенности входных данных модели можно устранять путем различных методов их устранения, таких как метод заполнения средним значением, подбором внутренних групп либо метод Hot Deck. В отличие от индуктивного подхода при создании моделей, в индуктивном подходе нет четкого критерия останова итеративного усложнения модели. Проверка модели на адекватность является одним из критериев завершения итерации жизненного цикла модели наряду с модульным и интеграционным тестированием программной реализации модели.

Список использованных источников

1. Кузьмин Е. А. Неопределенность и определенность в управлении организационно-экономическими системами / Е. А. Кузьмин. – Екатеринбург : Институт экономики УрО РАН, 2012. – 184 с.
2. Литвинов В. В. Создание блочных моделей систем и процессов с использованием метода группового учета аргументов / В. В. Литвинов, А. А. Задорожний // Математические машины и системы. – 2012. – № 2. – С. 107–116.
3. Глушков В. М. Синтез цифровых автоматов / В. М. Глушков. – М. : Физматгиз, 1962. – 476 с.
4. Томашевський В. М. Моделювання систем / В. М. Томашевський. – К. : Видавнича група ВНУ, 2007. – 352 с.