

УДК 638.51-77

Анализ и интерпретация больших объемов данных при управлении сетевыми объектами

Е.И. Сукач, М.Н. Гавриленко, А.А. Кончиц

Рассматривается способ обработки и анализа больших объемов данных, характеризующих сетевые объекты. Предлагается применение вероятностно-алгебраического моделирования, которое учитывает вероятностную природу объекта и существенно сокращает вычислительную сложность анализа объектов на этапах обработки, анализа и интерпретации.

Ключевые слова: вероятностное моделирование, надёжность, эффективность, большие данные, безопасность, многосвязные сетевые объекты, управление, вероятность, статистика.

A method for processing and analyzing large amounts of data characterizing network objects is considered. The application of probabilistic-algebraic modeling is proposed, which takes into account the probabilistic nature of the object and significantly reduces the computational complexity of the analysis of objects at the stages of processing, analysis and interpretation.

Keywords: probabilistic modeling, reliability, efficiency, big data, security, multiply connected network objects, management, probability, statistics.

Введение. Производственные, информационные, транспортные и другие сложные системы в ряде случаев имеют явно выраженную графовую структуру и включают множество компонентов, рост числа которых и усложнение связей между которыми сопровождается экспоненциальным ростом вычислений при попытке оценить их надёжность. Современный уровень развития техники, сопровождающий функционирование сетевых структур, позволяет фиксировать и накапливать информацию, отражающую динамику изменения вероятностных свойств компонентов выделенного класса объектов. Эти данные, полученные на протяжении длительного периода времени, имеют большой объем, отличаются многообразием, и в ряде случаев плохо структурированы. Рост вычислительной сложности алгоритмов и увеличение объема разнородных данных при исследовании реальных сетевых объектов, которые имеют тенденцию к увеличению размера, требуют применения современных методов и средств их автоматизации, позволяющих справиться с задачами сбора, систематизации, обработки и интерпретации имеющейся информации с целью повышения надёжности и эффективности функционирования исследуемых объектов, а также своевременного предупреждения возможных чрезвычайных ситуаций в работе оценки систем с потенциально опасными компонентами.

Для анализа свойств выделенного класса объектов традиционно используются методы статистического, кластерного и регрессионного анализа, методы распознавания образов, методы интеграции разнородных данных из разнообразных источников, искусственные нейронные сети, методы теории графов, сетевого анализа и оптимизации и др. Конечной целью их применения является выбор варианта устойчивого функционирования исследуемого объекта, при котором сетевая структура связана и параметры её надёжности находятся в заданных пределах.

Учитывая, что данные, характеризующие выделенный класс объектов, слишком велики, плохо структурированы и не форматированы, можно заключить, что для их анализа недостаточно аналитических методов и средств их автоматизации. Необходимы новые методы, технологии и инструментальные средства, расширяющие возможности работы с большими объемами данных [1], [2]. Определённой результативностью для изучения сетевых объектов с учётом изменения вероятностных свойств их структурных элементов и параметров внешней среды обладают методы компьютерного моделирования, позволяющие с учетом имеющейся информации построить адекватные модели и использовать их для поддержки принятия обоснованных решений с использованием вычислительной техники.

В статье рассматриваются возможности аппарата вероятностно-алгебраического моделирования (ВАЛМ) для обработки и анализа больших объемов данных, характеризующих

сетевые объекты различной структурной сложности [3]. Методы позволяют учесть вероятностную природу структурных составляющих исследуемых объектов, ускоряют процесс моделирования без потери точности и упрощения модели исследуемой системы.

Ввиду больших объемов информации и структурной сложности реальных объектов исследования оценка вероятностных свойств их функционирования с использованием существующих ручных технологий трудоёмка. Эта проблема решается посредством создания параметризованных компьютерных моделей сетевых объектов и разработки средств их реализации, включающих методы, методики и программное обеспечение, которые учитывают особенности организации исследуемых объектов, варианты их структурной организации, случайный характер происходящих в них процессов, особенности взаимодействия элементов между собой и с внешней средой.

Автоматизированная обработка исходных данных моделирования. Поскольку объектом исследования являются сети большой размерности с вероятностными параметрами их функционирования, то предлагается два этапа обработки и структурирования информации: обработка данных о структуре сетевого объекта; обработка данных натуральных экспериментов о функционировании объекта. На основе полученной информации выбирается схема формализации объекта и вид параметризованной модели.

Формирование графовой структуры модели сетевого объекта реализуется с использованием интеллектуального графического редактора поэтапно, а именно: декомпозиция объекта с учётом выбранного уровня детализации; определение вида графовой структуры (простая/структурно-сложная); понижение размерности графа модели; формирование вариантов разбиения сети на подструктуры n -полюсники; представление объекта в виде обобщённой графовой структуры со сложными вершинами (n -полюсниками).

Исходными данными для выполнения этой работы является детальная схема сетевого объекта с обозначенными структурными компонентами, которая используется в виде «подложки» и обеспечивает оперативность занесения информации. В составе программного инструментария реализован универсальный подход к формализации многосвязного сетевого объекта, обеспечивающий расчёт свойства сетевого объекта, как в случае схемы формализации «элементы-ребра», в которой варьируются вероятностные параметры рёбер графа, так и в случае схемы «элементы-вершины», предполагающей вероятностные изменения параметров вершин. Для структурно-сложных сетевых объектов реализована возможность эквивалентного преобразование одной схемы в другую. Для объектов простой графовой структуры, являющихся частным случаем схемы «элементы-вершины», реализована возможность одновременного рассмотрения вероятностных характеристик исследуемого свойства элементов системы и вероятностных функций взаимодействия между этими элементами. Эта особенность формализации отражает новизну подхода к формализации сетевых объектов и расширяет класс задач.

Понижение размерности графа реализуется в результате поэтапного анализа графа модели. На первом шаге сокращается число структурных элементов, влиянием которых можно пренебречь в процессе формирования оценки исследуемого свойства. На втором шаге выделяются фрагменты простой графовой структуры и реализуется алгоритм вероятностно-алгебраического моделирования с целью расчёта замещающего вероятностного значения исследуемого свойства фрагмента сетевого объекта. При этом исходная графовая структура модифицируется путём замещения каждого из найденных фрагментов одним элементом модели.

Далее для многосвязных сетевых объектов большой размерности с использованием специальных алгоритмов в составе графического редактора формируются возможные варианты представления графа модели в виде композиции подструктур n -полюсников ($n = 2,3,4$), выступающих в качестве элементов обобщённой модели. Вероятностные свойства подструктур оцениваются с использованием одной из методик, предназначенных для оценки свойств объектов ограниченной размерности. Один из вариантов используется для расчёта интегральных значений вероятностных свойства исследуемого объекта.

Оперативность построения графа модели с возможностью варьирования уровня детализации обеспечивается такими свойствами интеллектуального графического редактора, как иерархичность, масштабируемость, интерактивность, гибкость и прозрачность. Результатом выполнения первого этапа обработки является выбор параметризованной модели соответ-

ствующей структуре исследуемого объекта, которая может быть использована как «заготовка» при моделировании исследуемого свойства сетевого объекта и требуют лишь задания параметров элементов и настройки структуры.

В процессе реализации второго этапа (обработки данных наблюдения о функционировании сетевого объекта) обработки формируются параметры моделирования. При этом предполагается, что, во-первых, сетевые объекты являются неоднородными в смысле исследуемого свойства, и неоднородность эта может относиться как к узлам, так и к рёбрам, и, во-вторых, система наследует свойства своих компонентов.

Информация, полученная в процессе наблюдения, должна быть систематизирована и обработана. Из всего многообразия информации необходимо выбрать нужную и представить её в виде, удобном для анализа. Получение информации в своей содержательной, творческой части слабо поддаётся автоматизации. Однако, наличие технических средств индикации и сопряжения с базой данных (БД) в составе программного обеспечения обеспечивает формирование массивов численных значений физических параметров (отказы в работе с классификацией причин, пропускная способность, стоимость выполнения типовой операции и др.), характеризующих функционирование структурных составляющих $K = \{K_i\}, i = \overline{1, m}$ исследуемого объекта. Эта информация, систематизированная и помещённая в БД, обрабатывается с использованием средств интеграции разнородных данных и универсальных процедур статистических пакетов. В результате последовательной обработки данных наблюдения формируются параметры моделирования, определяющих вероятность несовместных состояний $S = \{S_j\}, j = \overline{0, n-1}$ компонентов, которые изменяются во времени и характеризуют исследуемое свойство сети:

$$P^{it} = (p_0^{it}, p_1^{it}, \dots, p_n^{it}), \sum_{j=0}^n p_j^{it} = 1, i = \overline{1, m}, t = \overline{1, T}. \quad (1)$$

Сформированные параметры фиксируются в БД программного инструментария и связывают геометрическое представление сетевых объектов, имеющих пространственную природу, с таблицами параметров, характеризующих свойства структурных составляющих этих объектов.

Автоматизированный анализ вероятностных свойств сетевых объектов. Эффективность обработки информации и анализа определяется поставленной задачей и имеющимися у исследователя средствами автоматизации. Как правило, на практике используются не самые современные методы обработки и анализа информации, позволяющие получить точное решение с учётом критерия выбора, а привычные инструментальные средства, которые имеются в наличии у исследователя. Следовательно, разработка и введение в практику современных и перспективных методов и средств их автоматизации позволит повысить уровень исследования и расширить класс решаемых задач.

Статическое вероятностно-алгебраическое моделирование. Статическое ВАЛМ реализуется с использованием одной из параметризованных моделей соответствующей выбранной схеме формализации на выбранном уровне детализации.

Наглядное представление элементов моделей, связей между ними и уровней вложенности функций, описывающих эти связи, реализуемое с использованием графического редактора, позволяет установить уровни иерархии функциональных связей и в дальнейшем, используя эффективные алгоритмы обхода вершин графа при машинной реализации методов, обеспечить замещение функциональных связей между элементами модели вероятностными вычислениями с использованием коэффициентов вероятностно-алгебраического моделирования.

Для сетевых объектов простой структурной организации из библиотеки параметризованных моделей программного обеспечения выбирается модель, не имеющая ограничений на число элементов и их состояний, и реализующая вероятностно-алгебраическое умножение векторов (1) согласно операциям, описывающим их взаимодействие. Для структурно-сложных сетей используется модель, реализующая методику сведения моделей с произвольным числом состояний к совокупности бинарных моделей. Для многосвязных сетевых объектов большой размерности реализуется расчётный метод вероятностно-алгебраического моделирования умножения подструктур n -полюсников.

Реализация многошагового процесса моделирования исследуемого сетевого объекта с использованием построенной модели сопровождается занесением в БД результатов моделирования в виде векторов вида

$$P^{net-t} = (p_0^{net-t}, p_1^{net-t}, \dots, p_n^{net-t}), \sum_{j=0}^n p_j^{net-t} = 1, i = \overline{1, m}, t = \overline{1, T}. \quad (2)$$

Результативность анализа многосвязных сетевых объектов различной структурной организации из разнообразных предметных областей, проводимого в рамках вероятностно-алгебраического аппарата, определяется автоматизацией проведения статического моделирования с использованием различных уровней вероятностно-алгебраического умножения (простые элементы/подструктуры n-полюсники) и специальных методик, обеспечивающих сокращение вычислительной сложности алгоритмов.

Упреждающее вероятностно-алгебраическое моделирование сетевого объекта.

Упреждающему вероятностно-алгебраическому моделированию должно предшествовать детальное изучение сетевого объекта. На основе полученной информации определяется состав технических средств и выбирается технология вероятностно-алгебраического моделирования.

Задачей упреждающего вероятностно-алгебраического моделирования сетевых объектов является выбор такой последовательности управляющих воздействий, при которой согласно полученным результатам вероятностно-алгебраического моделирования своевременно реализуются изменения в организации объекта. В этом смысле моделирование позволяет адаптировать сетевой объект к вероятностным изменениям свойств его компонентов. Вид управляющего воздействия выбирается с использованием одной из параметризованных моделей в составе программного инструментария, позволяющего рассмотреть объект на разных уровнях детализации, определив состав компонентов и указав семантику состояний этих компонентов.

Заключение. Согласованное использование различных стратегий ВАЛМ, включающих последовательное распределение компонентов сетевых объектов по иерархическим уровням и их интеграцию в единую систему сбора, обработки и анализа данных позволяет повысить эффективность управления объектами исследования и обеспечить решение следующих задач: организации и автоматизации оперативного мониторинга за состоянием исследуемого свойства структурных компонентов сетевых объектов; классификации состояний и формирования вероятностных оценок исследуемого свойства как отдельных компонентов, так и всего сетевого объекта; определения оптимального сочетания параметров исследуемого свойства структурных компонентов сетевого объекта, обеспечивающих заданный уровень надёжности и эффективности организации системы без изменения её структурной организации; оценки вариантов модификации структуры сетевого объекта по заданному критерию с использованием приемлемых по вычислительной сложности алгоритмов.

Литература

1. Гавриленко, М.Н. Разработка систем обработки и анализа данных с использованием инструментов BIG DATA / М.Н. Гавриленко // Творчество молодых, 2019 : сборник научных работ студентов, магистрантов и аспирантов : в 3 ч. / ред. кол.: Р.В. Бородич [и др.] ; Министерство образования Республики Беларусь, Гомельский государственный университет им. Ф. Скорины. – Гомель : ГГУ им. Ф. Скорины, 2019. – Ч. 1. – С. 214–217.

2. Гавриленко, М.Н. Анализ слабоструктурированных данных с использованием инструментов BIG DATA / М.Н. Гавриленко, А.А. Кончиц // Дни студенческой науки : материалы XLVIII студенческой научно-практической конференции, Гомель, 14–5 мая 2019 года : в 2 ч. / ред. кол.: Р.В. Бородич [и др.] ; Министерство образования Республики Беларусь, Гомельский государственный университет им. Ф. Скорины. – Гомель : ГГУ им. Ф. Скорины, 2019. – Ч. 1. – С. 52–53.

3. Сукач, Е.И. Вероятностно-алгебраическое моделирование сложных систем графовой структуры / Е.И. Сукач. – Гомель : ГГУ им. Ф. Скорины, 2012. – 224 с.