

ПОДХОДЫ К ИДЕНТИФИКАЦИИ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ВЕРХНЕГО СТРОЕНИЯ ЖЕЛЕЗНОДОРОЖНОГО ПУТИ

Владова А. Ю.

Лаборатория 41, Институт проблем управления им. В.А. Трапезникова РАН

Финансовый университет при Правительстве Российской Федерации

Москва, РФ

E-mail: {avladova}@mail.ru

Рассмотрены аналитический, статистический и обучающий (data science) подходы к оценке технического состояния верхнего строения железнодорожного пути как идентифицируемой системы. Представлены проблемы анализа технического состояния систем, достоинства и недостатки методов.

ВВЕДЕНИЕ

Верхнее строение железнодорожного пути — это система, в котором происходит взаимодействие между разнотипными признаками (инженерные, трибологические, климатические, геологические, гидрологические и др.) и формируются наблюдаемые сигналы. Система является динамической, в которой текущие значения наблюдаемых сигналов зависят не только от текущих, но и от более ранних значений внешних воздействий. Воздействия на систему в процессе эксплуатации доступны косвенной оценке по влиянию, оказываемую ими на выходные сигналы. Как следствие типового подхода к сбору информации выходные сигналы привязаны к дискретным моментам времени в общем случае распределенным неравномерно.

1. АНАЛИТИЧЕСКИЙ И СТАТИСТИЧЕСКИЙ ПОДХОДЫ

В теории идентификации связь между распределенным во времени внешним воздействием $u(t)$ и наблюдаемым сигналом $y(t)$ в линейной стационарной системе описывают импульсной реакцией (весовой функцией) $g(\tau)$:

$$y(t) = \int_0^{\infty} g(\tau)u(t - \tau)d\tau \quad (1)$$

Используя (1) можно провести оценку и прогнозирование наблюдаемого сигнала $y(t)$ если известно внешнее воздействие $u(t)$, что в большинстве случаев нереально. Всегда существуют помехи, влияющие на результат наблюдений: шум наблюдений — датчики подвергаются воздействию помех и дрейфу; неконтролируемые входы — в систему поступают сигналы, которые имеют характер входных, но не контролируются пользователем (например, сезонное промерзание рельс, подтопление). Линейная теория идентификации дает предположение, что влияние помех сводится к аддитивной компоненте $v(t)$ в наблюдаемом сигнале:

$$y(t) = \sum_{k=1}^{\infty} g(k)u(t - k) + v(t) \quad (2)$$

Наиболее характерной особенностью помехи является то, что ее значение заранее неизвестно. Однако информация о прошлых значениях помехи может использоваться для ее оценки в будущем. Для описания будущего поведения помехи принято использовать вероятностный подход. Однако реализация этого подхода требует [1]:

- формирования статистических гипотез о будущем поведении помех,
- задания различных функций плотности вероятности помех,
- анализ помех, являющихся смесью детерминированных и случайных компонент и пр.

Кроме того, для оценки технического состояния реальной системы в рамках теории идентификации часто необходим переход к системе нелинейных уравнений с запаздывающими аргументами [2], например, когда функция $g(t)$ заменяется на $g(t, t-k)$ и (2) записывается в виде:

$$y(t) = \sum_{s=-\infty}^{t-1} g(t, s)u(s) + v(t) \quad (3)$$

где $g(t, s)$ ($t=s, s+1, \dots$) - реакция на отклик в момент времени t на единичный входной импульс в момент s . С другой стороны, использование методов математической статистики в случае оценки технического состояния систем не всегда приводит к успеху. Одна из причин этого — концепция усреднения по выборке, приводящая к операциям над фиктивными величинами (например, средний пропущенный тоннаж, средний износ). Второй причиной является то, что погашение неточностей наблюдений при статистической обработке данных происходит в случае, если неточности являются результатом случайных погрешностей [3]. Если отклонения в данных носят неслучайный характер, то необходимо корректировать данные, либо собирать их заново. Третьей проблемой является то, что область применения статистических моделей ограничена ближайшей окрестностью наблюдаемых значений. Таким образом, проблемами анализа технического состояния систем являются [4]:

- проблема объема выборки или стохастической сходимости;
- проблема большого числа признаков (размерность модели);
- проблема многокомпонентного наблюдаемого сигнала. Один из способов решения – применение интегрирующей функции (например, функции полезности, эффективности или надежности);
- проблема нелинейности, когда одна из функций (2) или все три нелинейны.

II. DATA SCIENCE ПОДХОД

С другой стороны, современные системы мониторинга технического состояния объектов характеризуются тем, что емкости компьютерных хранилищ увеличились и стали дешевы, сенсоры уменьшились в размерах и способны отслеживать сигналы различной физической природы, а также передавать данные на значительные расстояния, а облачные вычисления на основе технологии анализа данных позволяют использовать возможности огромного количества машин для манипулирования этими данными. Технологии анализа данных при оценке технического состояния систем помогают решить следующие проблемы: выбор критерия селекции, списка признаков (размерности модели) и сложности модели – при синтезе модели методами машинного обучения сложный объект, имеющий сотни взаимосвязанных признаков, может быть сравнительно просто алгоритмически описан; выявление однотипных фрагментов многоаспектных взаимоотношений в данных. Эти фрагменты представляют собой закономерности, свойственные подвыборкам данных; тестирование полученных моделей: весь исследуемый массив данных разделяют на две неравные группы. Большая из них (обучающая) является исходным материалом для построения моделей, а меньшая представляет собой тестовую группу, на которой проверяются полученные модели. Критерием, по которому оценивается модель, является разность в точности оценки между группами. Один из подходов в рамках применения технологии Data Science предлагает статья [6], посвященная оценке воздействия деформированных колес на железнодорожный путь путем полноволнового численного моделирования на суперкомпьютерных системах. Для специалистов в сфере машинного обучения статья интересна тем, что перечисляет целый ряд факторов, влияющих на зависимую переменную, например, степень поврежденности колесной пары, наличие исходных дефектов в рельсе, методика укладки шпал, влияние пустот под шпалами и др. Авторы статьи [7] предлагают проводить классификацию режима работы устройств железнодорожной автоматизации и телемеханики (в частности, электри-

ческой рельсовой цепи, работающей в трех режимах) на основе логистической регрессии или метода опорных векторов с гауссовым ядром. При этом вектор признаков формируют по параметрам электрических сигналов на входе и выходе устройства. В статье [8] проведен многофакторный анализ и кластеризация данных по отступлениям и неисправностям пути, архивам капитальных ремонтов, рельсовым книгам, активах земляного полотна, ведомостям дефектных рельсах, выполнено статистическое исследование ряда признаков и установлена нелинейная взаимосвязь между ними.

III. ЗАКЛЮЧЕНИЕ

Анализ литературы показал, что сформированы подходы на основе статистических, вероятностных и обучающих методов по отдельным типам данных. Современные технологии анализа данных позволяют интегрировать математические и статистические методы с программированием и способами хранения данных. Для решения проблемы идентификации технического состояния верхнего строения пути предложено интегрировать данные различной физической природы (пропущенный тоннаж, паспортные характеристики рельс, особенности плана и профиля пути и др.).

IV. СПИСОК ЛИТЕРАТУРЫ

1. Льюнг, Л. Идентификация систем. Теория для пользователя /Л. Льюнг // М.: Наука. Гл. ред. физ.-мат. лит., 1991. - 432 с.
2. Принятие решений на основе самоорганизации /Ивахненко [и др.]. - Изд-во: М.: Советское радио, 1976. - 280 с.
3. Венецкий, И. Г. Основные математико-статистические понятия и формулы в экономическом анализе / И. Г. Венецкий, В. И. Венецкая // М.: Статистика, 1979. - 447 с.
4. Нейлор, Т. Машинные имитационные эксперименты с моделями экономических систем / Т. Нейлор - Изд-во: Мир, Москва, 1975. - 504 с.
5. Федосеев, А. А. Технология data mining в задачах прогнозирования развития транспортной инфраструктуры [Электронный ресурс] / Современные проблемы науки и образования. - 2013. № 1. - Режим доступа: <http://www.science-education.ru/ru/article/view?id=8153>. - Дата доступа: 06.06.2021.
6. Фаворская, А. В. Разработка математических моделей, численных методов и расчетных программ для выявления дефектов элементов системы «колесо-рельс» / Фаворская [и др.]. - Бюллетень Объединенного ученого совета ОАО РЖД. № 1, 2018. с. 49–63.
7. Prisukhina, I. V. Machine state classification of electric track circuit by means of support vector machine / I. V. Prisukhina, D. V. Borisenko // Omsk Scientific Bulletin, 2018. с. 126–130.
8. Владова, А. Ю. Применение технологии анализа данных для идентификации технического состояния верхнего строения пути /Труды 14-й Международной конференции «Управление развитием крупномасштабных систем». М.: ИПУ РАН, 2021 (в печати).