

8. Ольховиков С. Э., Петренева Е. А. Проблемы логистики и хранения зерновых грузов и пути их решения // Вестник транспорта Поволжья. 2023. № 6(102). С. 78–86. EDN HNCCZL.
9. Гладунов В. А., Бондаренко Е. М. Выбор оптимальных логистических схем поставок товаров из Китая в Российскую Федерацию при использовании контейнеров // Вестник Сибирского государственного университета путей сообщения. 2023. № 1(64). С. 15–23. DOI 10.52170/1815–9265_2023_64_15. EDN KHWKOS.
10. Формирование узловых мультимодальных транспортно–логистических центров / С. Э. Ольховиков, Е. А. Петренева, И. Н. Кагадий, О. Б. Шерстобитова // Вестник Ростовского государственного университета путей сообщения. 2023. № 1 (89). С. 106–118. DOI 10.46973/0201–727X_2023_1_106. EDN SGTOBW.

УДК 656.259.12

© 2024 Е. М. Тарасов, И. С. Бредун, В. А. Надежкин, С. А. Надежкина

ПРИНЦИП ФОРМИРОВАНИЯ БАЗ ДАННЫХ И СИНТЕЗ КЛАССИФИКАТОРА СОСТОЯНИЙ УСТРОЙСТВ ЖАТ

В статье рассматриваются вопросы синтеза системы распознавания для классификации состояний первичных преобразователей информации, представленных длиной линии и описываемых уравнениями состояний в виде телеграфных уравнений. В первой части работы представлены вопросы формирования баз данных (априорных информативных признаков распознавания), принцип формирования образов состояний и алгоритм циклического функционирования системы при распознавании. Сформулированы требования по созданию (разбиению) пространства выборки информативных признаков на обучающие, тестовые и контрольные выборки. Рассмотрены этапы синтеза классификатора по различным критериям и представлен итерационный процесс качественного изменения первоначального варианта структурной сложности классификатора. В статье рассматриваются имеющиеся методы предиктивной аналитики, включая применение математических алгоритмов и скоринговых моделей, позволяющих своевременно выявлять определенные закономерности и на основе этого с высокой точностью прогнозировать дальнейшее поведение данных. Особое внимание в работе авторы уделяют возможности создания адаптивных систем, которые способны оперативно реагировать на возникшие изменения и результативно анализировать текущие события с целью прогнозирования будущих показателей.

Ключевые слова: прогностическая модель, математические алгоритмы, пропускная способность, передача данных, регрессивный анализ, распознавание образов, предиктивная аналитика, составление прогнозов, информативные признаки, объем данных.

Введение. Постоянное расширение списка информативных признаков, развитие технологий передачи, расширение пользовательской базы приводит к постоянной потребности в повышении скорости передачи данных на уровне доступа. В условиях невозможности увеличения потока данных вследствие ограниченности сетевой инфраструктуры необходимым инструментом для решения проблемы становится использование буферизации данных систем предиктивной аналитики, позволяющей минимизировать влияние временных задержек за счет накопления и хранения информации в промежуточных хранилищах. В свою очередь, буферизация данных способствует созданию адаптивных систем, которые могут своевременно обрабатывать массивы данных и проводить сложные вычисления.

Предикативная или прогностическая аналитика (Predictiveanalytics) – это, прежде всего, множество методов статистики, анализа данных, которые используются для анализа текущих данных/событий для прогноза данных/событий в будущем. Она представляет собой серьезный набор методов статистики и прогноза информации для интерпретации текущих данных и событий с задачей прогнозирования будущих высокотехнологических результатов.

Материалы и методы

Наиболее известный способ использования прогностической аналитики – применение скоринговых моделей для оценки работоспособности устройств. «Любая скоринговая модель строится на исторических данных, и если в прошлом, какая-либо группа элементов устройства была уличена в неисправности, а новая техническая база схожа с предыдущей, то вероятность возникновения такого же рода неисправностей достаточно высока» [1].

В основе предиктивной аналитики лежит использование математических алгоритмов для выявления и объяснения закономерностей в собираемых данных для составления прогнозов будущего поведения этих данных. Конкретные методы предиктивной аналитики, используемые в разработке систем, могут различаться в зависимости от конкретной задачи и применяемого инструментария. Однако, в целом, для решения задач предсказания и оптимизации в разработке систем часто используются следующие методы:

1. Алгоритмы распознавания образов.

Распознавание образов (РО), алгоритм которого приведен ниже на рис. 1, заключается «в поиске корреляции между зависимой переменной (временем выполнения задачи) и одной или несколькими независимыми переменными» [2].



Рис. 1. Алгоритм поведения системы при распознавании образов состояний

Для применения РО в разработке систем предиктивной аналитики необходимо собрать данные о зависимой переменной (в данном случае – время выполнения задачи) и одной или нескольких независимых переменных (например, сложность задачи, количество строк кода, используемые ресурсы и т. д.) [3].

Далее, данные необходимо подготовить: удалить выбросы и пропущенные значения, масштабировать данные при необходимости. Для того, чтобы избежать стадии переобучения модели, данные разбиваются на обучающую и тестовую выборки в соотношении 70/30 или 80/20, а также применяют кросс – проверочную выборку в процентном соотношении – 15 % – от общего объема выборки.

Затем выбирается тип решающей функции (РФ) в зависимости от данных и задачи. Например, для простой зависимости между временем выполнения задачи и одной независимой переменной может быть использована линейная РФ.

$$d(X) = f(x_1, x_2, \dots, x_n),$$

где x_1, x_2, \dots, x_n – переменные, исходные данные, аргументы полинома РФ;

X – образ состояний.

Если же зависимость более сложная, может быть применена полиномиальная или другая модель.

В последствии на обучающей выборке модель обучается, т. е. находятся коэффициенты РФ, определяющие зависимость между зависимой и независимыми переменными [4]. После обучения модель проверяется на тестовой выборке, где оценивается ее точность.

Наконец, применяя обученную модель, можно предсказывать время выполнения задачи при заданных значениях независимых переменных. Например, для задачи с определенным количеством строк кода и использованных ресурсов можно предсказать период времени выполнения на основе обученной модели [5].

Применение РФ в разработке систем предиктивной аналитики может быть полезным для планирования проектов, оптимизации производительности элементов систем, выявления проблемных мест в коде и других профессиональных задач. Кроме того, РО является одним из наиболее распространенных методов построения предикативных моделей.

Объем информации напрямую зависит от степени сортировки данных или уже принятых на ее основе решений. Так называемые сырые данные занимают достаточную площадь информационного пространства в общем потоке информации, когда «ответ», будь то предсказывающая или предписывающая информация, занимает значительно меньше места в общем потоке. С переходом на новые методы и системы передачи информации, отказываются новые возможности в качестве и информативности передаваемой информации.

«Во время переобучения (overfitting) негативное явление заключается в том, что функции $d(X)$ вырабатывает предсказания, которые слишком близко или точно соответствуют конкретному набору данных и поэтому не подходят для применения алгоритма к дополнительным данным или будущим наблюдениям» [6].

«Во время недообучения (underfitting) негативное явление заключается в том, что функции $d(X)$ не обеспечивает достаточно малой величины средней ошибки на обучающей выборке» [6]. Недообучение возникает при использовании недостаточно сложных моделей.



Рис. 2. Пирамида объема данных

В настоящее время широкое распространение получили 2 вида синтеза систем схем электрических устройств:

- структурный синтез;
- параметрический синтез.

При параметрическом синтезе структурная схема классификатора известна, и требуется определить параметры элементов схемы классификатора, отвечающие требованиям по функционированию.

Классически синтез рельсовых цепей выполняется в три этапа.

На первом этапе решается задача определения максимальной длины рельсовой линии и оптимальных входных сопротивлений по концам РЛ при заданных исходных первичных параметрах.

На втором этапе синтеза решается задача обеспечения стабильности параметров входных сопротивлений при изменении первичных параметров РЛ в допущенных пределах.

На третьем этапе с помощью общих методов и результатов синтеза разрабатывается схема рельсовой цепи.

При синтезе классификатора необходимо решать ряд проектных процедур, представленных на рис. 3.

Как отмечено выше, при параметрическом синтезе (структура классификатора известна) определяются числовые значения коэффициентов РФ, вид и сложность РФ в диапазоне возможного изменения внешних воздействий и первичных параметров. «При этом становится целью достижения некоторого экстремума целевой функции, то имеет место процедура оптимизации (параметрической оптимизации)» [7].

«В процессе оптимизации определяются номинальные параметры элементов измерительного канала» [8].

Процедура создания оптимального классификатора предполагает взаимосвязь синтеза и анализа, пример взаимосвязи представлен на рис. 4.

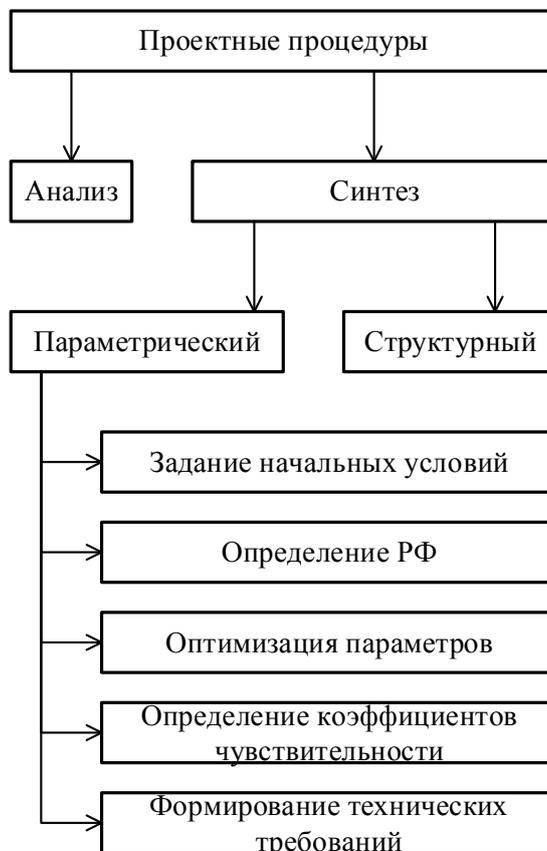


Рис. 3. Классификация проектных решений

Как правило, «по результатам анализа организуется итерационный процесс качественного измерения первоначального варианта классификатора измерениям параметров элементов согласующих устройств Z_o и Z_n . Этот процесс формализуется представлением как решение задачи параметрической оптимизации» [9].

В случае невыполнения требований по качеству классификации изменениям параметров элементов осуществляется возврат в начало процедуры синтеза и изменяется структура (вид или сложность решающей функции классификатора). Если и после этого заданное качество классификации не достигнуто, то корректируются начальные условия синтеза (изменяется длина РЛ контрольного участка или величина начальной проводимости изоляции).

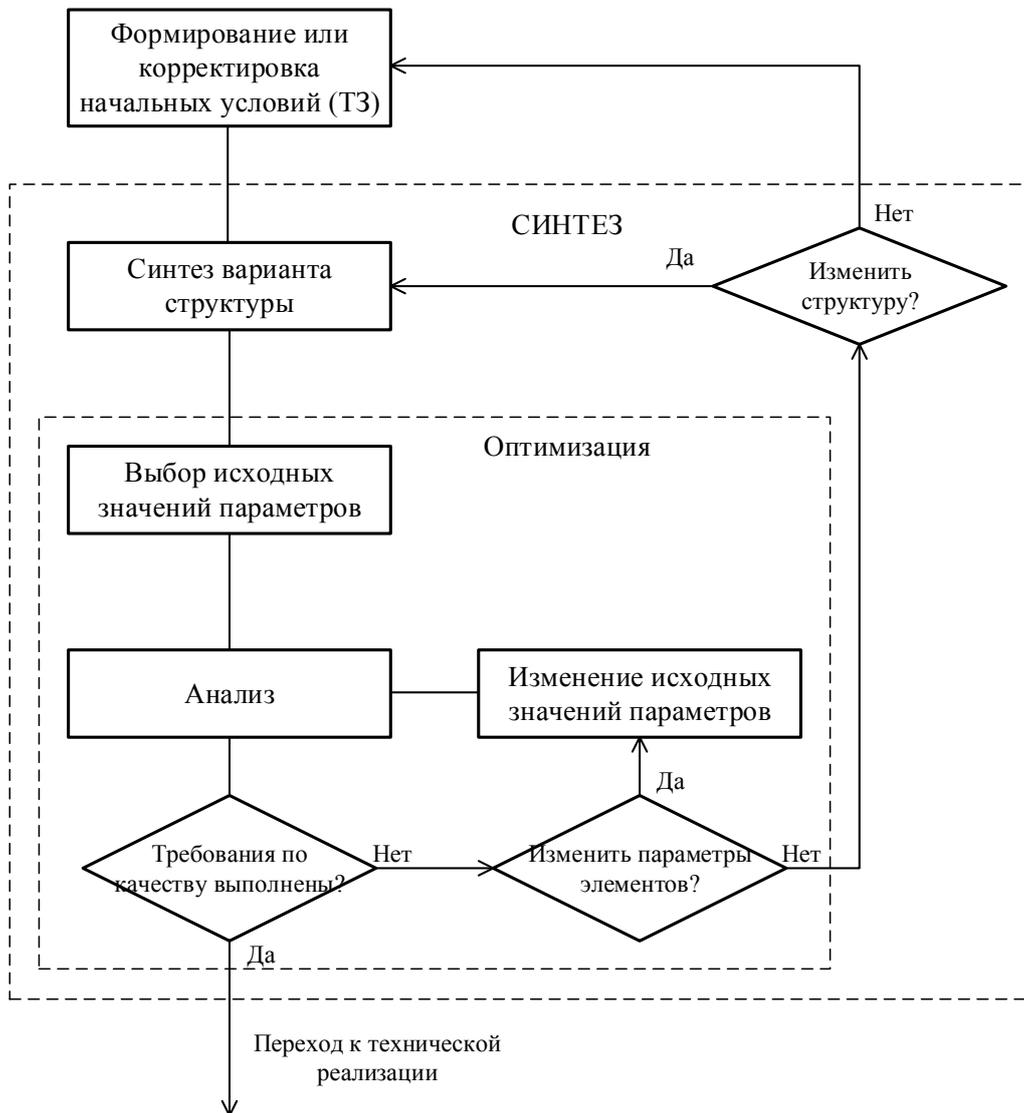


Рис. 4. Типовая схема этапов проектирования

Заключение. Таким образом, буферизация данных представляет собой эффективный инструмент для ведения предиктивной аналитики в условиях ограниченного потока данных. С учетом невозможности увеличения объема поступающей диагностической информации применение буферов позволяет обработать входящий поток данных для более точного и своевременного прогнозирования.

Кроме того, надежная организация систем буферизации в комплексном сочетании с продвинутыми алгоритмами анализа данных дает условия для формирования адаптивных решений, позволяющих в моменте реагировать на любые появляющиеся изменения как в массиве данных, так и непосредственно в условиях их обработки. Именно поэтому применение интеграции буферизации данных и современных аналитических подходов дает возможность эффективно использовать ресурсы и с высокой точностью оптимизировать технологические процессы.

СПИСОК ЛИТЕРАТУРЫ

- 1 Тарасова А. Е. Исследование потенциальных возможностей классификаторов состояний рельсовых линий // Наука и образование транспорту. 2018. № 1. С. 223–226.
- 2 Дмитренко И. Е. Техническая диагностика и автоконтроль систем железнодорожной автоматики и телемеханики. М.: Транспорт, 1986. 144 с.
- 3 Тарасов Е. М. Инвариантные классификаторы состояний рельсовых линий для систем интервального управления движением поездов: специальность 05.13.05 «Элементы и устройства вычислительной техники и систем управления»: автореферат диссертации на соискание ученой степени доктора технических наук / Тарасов Евгений Михайлович. Самара, 2004. 32 с.

- 4 Сапожников В. В., Сапожников Вл. В., Шаманов В. И. Надежность систем железнодорожной автоматики, телемеханики и связи. М.: Маршрут, 2003. 263 с.
- 5 Тарасов Е. М., Железнов Д. В., Васин Н. Н., Тарасова А. Е. Обеспечение инвариантности к возмущающим воздействиям в рельсовых линиях // Инженерные технологии и системы. 2019. Т. 29. № 2. С. 152–168.
- 6 Потапов А. С. Распознавание образов и машинное восприятие: общий подход на основе принципа минимальной длины описания. СПб.: Политехника, 2007. 548 с.
- 7 Тарасов Е. М., Тарасова А. Е., Надежкин В. А. К вопросу автоматизации технической диагностики и мониторинга // Наука и образование транспорту. 2022. № 1. С. 362–364.
- 8 Тарасова А. Е., Васин Н. Н. Формирование обобщенных решающих функций при распознавании и классификации состояний ЖАТ // Наука и образование транспорту. 2022. № 1. С. 360–362.
- 9 Тарасова А. Е. Принципы построения классификатора состояний рельсовых линий, инвариантного к изменению сопротивления рельсовых линий // Вестник СамГУПС. 2020. № 2(48). С. 82–88.