

УДК 004.04

ПРИМЕНЕНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ОБРАБОТКИ МЕТРИК ИСПОЛНЕНИЯ ЦОАСУ

Бакшевникова А.С., Привалов М.В.

*Донецкий национальный технический университет,
кафедра автоматизированных систем управления
bakshevnikova@gmail.com*

В статье описана математическая постановка задачи планирования и прогнозирования исполнения в целеориентированной АСУ (ЦОАСУ), рассмотрены некоторые известные методы анализа данных, применимые к таким задачам. Определены основные преимущества и недостатки использования средств data mining для анализа метрик исполнения.

Постановка проблемы

Целеориентированный подход к проектированию предполагает изучение и анализ целей создания автоматизированной системы для эффективной её реализации и унификации подхода к проектированию АСУ, а также последующего упрощения процесса интеграции систем. В основе работы целеориентированных систем лежит принцип непрерывного многоуровневого управления: их архитектура включает управляющие и управляемые блоки, стратегия работы которых изменяется во времени, то есть постоянно корректируется с учётом модификации целей, данных о внутреннем состоянии системы и об изменениях во внешней среде. Для принятия решений по корректировке управляющих воздействий необходимы актуальные сведения о процессе выполнения целей, называемые метриками исполнения.

В настоящее время не существует единого подхода к анализу данных метрик исполнения и алгоритмов их обработки. В данной статье описана задача прогнозирования метрик и рассмотрены возможные методы её решения.

Организация управления в ЦОАСУ

Работа ЦОАСУ в целом и каждой из отдельных её составляющих представляет собой непрерывный цикл управления, состоящий из четырёх этапов: наблюдение, ориентирование, решение, действие [1]. Подобный цикл характерен для любого процесса управления в кибернетике. Стадия наблюдения соответствует сбору оперативной информации о работе системы и состоянии внешней среды; стадия ориентирования – определению проблемных ситуаций и несоответствий плану выполнения; стадия решения – планированию дальнейших действий по решению проблемы, и, наконец, стадия действия соответствует всем операциям системы, выполняемым в соответствии с принятым решением. Сведения о выполнении операций затем передаются подсистеме анализа в качестве наблюдений, что обеспечивает обратную связь.

Каждой стадии цикла управления в ЦОАСУ соответствует отдельный функциональный блок. Принятие решений осуществляется с помощью систем

поддержки принятия решений различного уровня автоматизации: от ручных до полностью автоматических.

Работу решающих элементов ЦОАСУ можно представить в виде параметрической программы управления, направленной на поддержание заданных показателей исполнения целей (скорости выполнения, уровня потребления ресурсов и т.д.), которая постоянно корректируется в процессе работы. Система осуществляет управление неизбежно с некоторыми ошибками [2], информация о которых должна поступать в подсистему анализа для поддержания обратной связи. Функция анализирующего блока заключается в обработке этой информации, называемой метриками исполнения, для последующего построения прогнозов изменения наблюдаемых параметров. На основе таких прогнозов могут быть выявлены возможные несоответствия фактических результатов исполнения запланированным. В этом случае решающие элементы системы должны будут скорректировать управляющие воздействия с учётом обнаруженных несоответствий. Таким образом, анализ метрик исполнения служит основным инструментом для оперативного и стратегического планирования и прогнозирования достижения целей.

Анализ и прогнозирование исполнения

Планирование представляет собой ориентированную на будущее деятельность по принятию решений, которая включает назначение и выбор вариантов поведения. Оно связано с достижением желаемых будущих состояний и обращается к вопросу, что должно быть сделано и в какой степени обусловлена эта предполагаемая деятельность. Сложность процесса планирования и принятия решений заключается в достижении компромисса между долгосрочными (стратегическими) и краткосрочными (тактическими) целями и включает рациональный выбор альтернатив и средств для достижения целей [3]. Основой для планирования служат исторические данные, по которым строится прогноз исполнения.

Задачи прогнозирования в ЦОАСУ сводятся к двум основным типам: прогнозирование времени достижения цели и прогнозирование значений параметров в заданные моменты времени. Математическая модель анализа метрик может быть представлена следующим образом: пусть

$P = \{p_1, p_2, \dots, p_N\}$ – наблюдаемые параметры системы;

N – количество параметров;

$Y_{1..N}^P = \{Y^{p1}, Y^{p2}, \dots, Y^{pN}\}$ – набор временных рядов, описывающих изменение значений параметров P , где

$Y_{1..ni}^{pi} = \{y_{1'}^{pi}, y_{2'}^{pi}, \dots, y_{ni'}^{pi}\}$ – набор последовательных измерений значений P_i в моменты времени от 1 до n_i (в общем случае n_i различно для всех P_i);

$V^P = \{V^{p1}, V^{p2}, \dots, V^{pN}\}$ – целевые значения параметров, где

$V^{pi} = \{V_{1'}^{pi}, V_{2'}^{pi}, \dots, V_{ni+h}^{pi}\}$ – целевые значения параметра P_i ;

$t_g = \{t_g^{p1}, t_g^{p2}, \dots, t_g^{pN}\}$ – целевое время достижения V^P ;

$t_r = \{t_r^{p1}, t_r^{p2}, \dots, t_r^{pN}\}$ – реальное время достижения V^P ;

$t_p = \{t_p^{p1}, t_p^{p2}, \dots, t_p^{pN}\}$ – прогнозируемое время достижения V^P .

$Y_{ni+1..hi}^{pi} = \{y_{ni+1}^{pi}, y_{ni+2}^{pi}, \dots, y_{ni+hi}^{pi}\}$ – прогноз для параметра P_i на h_i периодов (в общем

случае h_i различно для каждого P_i).

Управляющая система стремится уменьшить разницу между прогнозируемыми и целевыми значениями параметров:

$$E_{\text{val}} = \sum_{i=1}^N \sum_{j=ni+1}^{hi} (V_{hi}^{Pi} - y_{hi}^{Pi})^2 \text{ в случае прогнозирования значений параметров или}$$

$$E_{\text{time}} = \sum_{i=1}^N (t_p^{Pi} - t_g^{Pi})^2 \text{ в случае прогнозирования времени достижения цели.}$$

Критерием оптимизации может служить минимизация суммарной квадратичной ошибки управления E :

$$E_{\text{val}} = \sum_{i=1}^N \sum_{j=ni+1}^{hi} (V_{hi}^{Pi} - y_{hi}^{Pi})^2 \text{ или } E_{\text{time}} = \sum_{i=1}^N (t_g^{Pi} - t_r^{Pi})^2.$$

На рис. 1 показан прогресс изменения наблюдаемого параметра P_i (задача прогнозирования значения параметра): в момент времени t_1 фактическое значение Y_{P1} ниже запланированного VP_1 , в момент t_3 – выше запланированного, в момент t_2 значения совпадают.

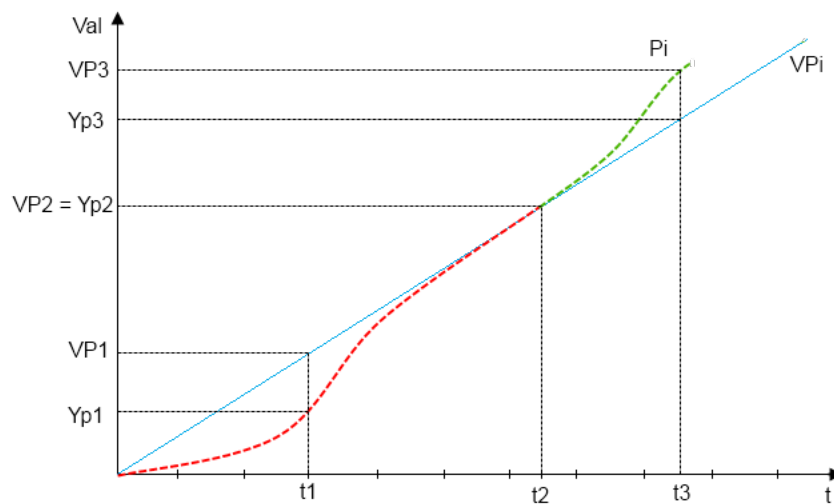


Рисунок 1. Запланированные и фактические результаты исполнения

Блок наблюдения осуществляет непрерывную регистрацию изменений внутреннего состояния системы и внешней среды, то есть метрики исполнения представляют собой информацию, непрерывно накапливаемую в системе и характеризующую её состояние в различные моменты времени. Это означает, что они могут быть представлены в виде временных рядов наблюдаемых параметров, на поддержание которых ориентируется решающая структура. Следовательно, для принятия решений на основе таких данных применимы методы анализа и прогнозирования временных рядов.

Временной ряд (ряд динамики) – это последовательность упорядоченных во времени числовых показателей, характеризующих уровень состояния и изменения изучаемого явления. Временной ряд включает два обязательных элемента: время и уровень ряда – конкретное значение показателя [4], в данном случае значение метрики. Прогнозирование временного ряда заключается в поиске наилучшей в некотором смысле оценки $y[N+h]$ ненаблюдаемой величины $y[N+h]$ (или оценки t_h времени, необходимого для достижения этой величиной заданного значения) по наблюдениям $y[n]$, $n=0, N$, то

есть как функции этих наблюдений, где

$y[N+h]$ – ненаблюдаемый уровень ряда, соответствующий моменту времени t_{N+h} ;

t_i – момент времени i ;

y – исследуемый временной ряд;

N – количество уровней временного ряда y [4].

Определение степени зависимости переменных и предсказание значений наблюдаемых величин являются одной из основных задач анализа временных рядов.

Методы и средства прогнозирования

Задача прогнозирования успешно решается методами математической статистики и средствами интеллектуального анализа данных, среди которых наиболее часто используются нейронные сети (НС) и деревья решений.

Прогнозирование в математической статистике представляет собой подбор аналитического выражения $y = F(x)$, наиболее точно описывающего тренд - зависимость уровней ряда от времени, и последующую экстраполяцию ряда для получения предсказанных значений. Ключевым этапом является выбор модели регрессии - вида уравнения, описывающего зависимость. Простейший вид модели - линейная регрессия. На практике также часто используются аппроксимирующие многочлены. Подбор параметров регрессии выполняется, как правило, методом наименьших квадратов. Главным недостатком использования статистических подходов является проблема выбора модели регрессии: недостаточно сложные модели не смогут обеспечить необходимую точность, а модели, обладающие избыточной сложностью, могут оказаться переобученными. Так как реальные автоматизированные системы, как правило, являются сложными, то применение регрессионных моделей затруднено. Альтернативой являются методы интеллектуального анализа данных, не требующие априорно известной аналитической модели.

Нейронные сети в настоящее время являются одним из наиболее популярных и стремительно развивающихся направлений исследований в области искусственного интеллекта. Задачи управления и регулирования с предсказанием хорошо изучены и успешно решаются с помощью НС [5].

Задача прогнозирования для НС сводится к задаче аппроксимации - поиска такой зависимости, которая наилучшим образом отражала бы изменение временного ряда во времени. Для прогнозирования применяются однослойные линейные сети с временной задержкой (TDNN), многослойные полносвязные сети прямого распространения (многослойные персептроны), сети на основе радиально-базисных функций (RBF-сети), обобщённо-регрессионные (GRNN-сети), некоторые специфические виды сетей с частичными связями специальной структуры, а также нейросети нечёткого вывода [6].

Однослойная линейная сеть имеет наиболее понятную и простую структуру, но требует значительного времени и объёма выборки для обучения. Введение временной задержки приводит к тому, что обработка входных значений зависит от выходных значений на предыдущем шаге работы.

Многослойный персептрон - хорошо изученная модель НС с широкими возможностями, применимая для задачи аппроксимации. Для обучения, как правило,

используется метод обратного распространения ошибки. Важным вопросом при использовании многослойных перцептронов является определение количества скрытых слоёв и числа элементов в них. Оптимальный выбор архитектуры такой сети зависит от конкретной задачи.

Обобщённо-регрессионная сеть, предложенная Шпехтом [7], предназначена для построения обобщённых (линейных и нелинейных) регрессий. Использование такой сети не требует априорных знаний о виде регрессионной зависимости. Обучение GRNN-сети осуществляется особым образом и не требует итерационной подстройки весов. Более быстрый и простой процесс обучения делает возможным применение GRNN-сети для работы в реальном времени.

RBF-сети основаны на использовании скрытого шаблонного слоя нейронов с радиально-базисными активационными функциями (гауссова, мультиквадратичная, функция Коши и т.д.). Выходной слой образуют нейроны с линейной функцией активации. RBF-сети также обладают высокой скоростью обучения.

Важным свойством НС, которое обеспечивает значительное преимущество их использования, является адаптивность: сеть подстраивает синаптические веса под изменения окружающей среды. Для работы в нестационарной среде, где статистика изменяется с течением времени, могут быть созданы НС, изменяющие синаптические веса в реальном времени, что делает возможным создание систем адаптивного прогнозирования [5].

Деревья решений – ещё один метод интеллектуального анализа данных, применяющийся для задач регрессии и классификации и кластеризации. Задача регрессии сводится к выявлению зависимости целевой переменной от независимых (входных) переменных. К этому же классу относятся задачи численного прогнозирования.

В задачах прогнозирования применяются регрессионные деревья решений, оперирующие численными значениями атрибутов. Для их построения используются алгоритмы CART, SLIQ, SPRINT.

Основной сложностью в процессе конструирования дерева является выбор критерия остановки разбиения или отсечения ветвей. Критериями остановки, как правило, служит максимальное количество уровней дерева (глубина дерева) или минимальное число примеров, которые содержит конечный узел. Алгоритмов сокращения деревьев достаточно много (cost-complexity, reduced error, MDL-based pruning [8]).

Существенное преимущество применения деревьев решений для прогнозирования заключается в интуитивно понятном представлении логики работы решающего аппарата. Ещё одним плюсом в сравнении с нейронными сетями является простой и более быстрый процесс обучения: для построения дерева достаточно задать набор всех имеющихся входных атрибутов, алгоритм построения выберет наиболее значимые из них.

Выводы

Для определения методов анализа исполнения выполнена математическая постановка задачи планирования и прогнозирования на основе метрик. Выяснено, что метрики исполнения представляют собой параметры системы, позволяющие оценить

прогресс достижения цели и проанализировать процесс выполнения с целью выявления несоответствий и проблемных ситуаций. Показано, что сведения об изменении таких параметров могут быть формализованы в виде временных рядов и для их анализа применимы известные средства прогнозирования: методы математической статистики, нейронные сети и деревья решений.

В ходе анализа перечисленных методов прогнозирования определены наиболее существенные преимущества и недостатки средств интеллектуального анализа данных в применении к рассматриваемой задаче.

Список источников

- [1] McElman, C. From Buttons to Bits – Achieving Level 3 Integration. Paper 4383, APCOM 2009 conference, Vancouver BC, Oct. 2009. /C. McElman, S. Seroukhov// Интернет-ресурс. – Режим доступа: [www/URL: http://www.mmsi.com/files/download/TechPapers/McElman_Seroukhov_APCOM_2009.pdf](http://www.mmsi.com/files/download/TechPapers/McElman_Seroukhov_APCOM_2009.pdf). – Загл. с экрана.
- [2] Бесекерский В.А., Попов Е.П. Теория систем автоматического управления. СПб. : Профессия, 2003 г. – 751 с.
- [3] Саати Т., Керис К. Аналитическое планирование. Организация систем: Пер. с англ. М. : Радио и связь, 1991 г. – 224 с.
- [4] Чураков Е.П. Прогнозирование эконометрических временных рядов: Учеб. пособие / Е.П.Чураков. М. : Финансы и статистика, 2008 г. – 208 с.
- [5] Хайкин, Саймон. Нейронные сети: полный курс, 2-е издание. : Пер. с англ. М. : Издательский дом “Вильямс”, 2006. 1104 с.
- [6] Рутковская Д., Пилиньский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечёткие системы: Пер. с польск. И.Д.Рудинского. М. : Горячая линия-Телеком, 2006. – 452 с.
- [7] Specht D.F. A General Regression Neural Network // IEEE Trans. on Neural Networks. – 1991. – 2. – № 6.
- [8] Mehta M., Rissanen J., Agrawal R. MDL-based Decision Tree Pruning // Proceedings of the 1st International Conference on Knowledge Discovery and Data Mining. Montreal, Canada. – 1995. – P. 216–221.