

дент в сервис-деске. К инциденту прикрепляются соответствующие атрибуты (например, номер автомобиля, масса автомобиля, время в пути и т. д.), а также (при возможности) подтверждающие инцидент кадры с видеокамер. Помимо этого диспетчер может просмотреть видеоролик зафиксированного события, который берется из архива системы видеонаблюдения. Таким образом, диспетчер, заметив новый инцидент, может сразу понять причину его создания и принять соответствующее решение.

В заключение следует отметить, что интеграция системы видеоаналитики с датчиками и информационными системами предприятия в единую комплексную систему видеонаблюдения позволяет контролировать всю цепочку технологического процесса в реальном масштабе времени и быстро принимать решения в зависимости от типа и важности зафиксированных событий и инцидентов.

S. S. Golyak, S. A. Gordeev, I. A. Goroshkov, G. V. Razumovsky

INTEGRATED VIDEO SURVEILLANCE SYSTEM FOR TECHNOLOGICAL PROCESS MONITORING

This paper is focused on a monitoring of technological processes via integrated video surveillance system. Authors consider software architecture issues, problems of interaction with other enterprise information systems and ways of usage video surveillance system for technological process monitoring.

Video Surveillance System (CCTV), Video Content Analytics (VCA), Event Processing Server, Video Controller

УДК 512.2

Н. А. Жукова, И. С. Соколов, А. В. Экало

Метод нечеткого сравнения телеметрических параметров на основе алгоритмов интеллектуального анализа данных

Предлагается метод нечеткого сравнения медленно меняющихся телеметрических параметров на основе символического представления с использованием взвешенного редакционного расстояния. Применение редакционного расстояния позволяет сравнивать параметры, смещенные друг относительно друга как по оси времени, так и по оси значений, а также не учитывать несущественные отклонения в поведении параметров, вызванные влиянием внешних факторов.

Интеллектуальный анализ данных, символическое представление, телеметрические параметры

Содержание современной телеметрии составляет широкий круг проблем, связанных с получением, преобразованием, передачей и обработкой телеметрической информации, используемой при управлении удаленными объектами, определении их состояния или при изучении физических процессов в местах, где непосредственное присутствие наблюдателя затруднено или невозможно. Под телеметрируемым параметром (ТПП) понимается показатель физического процесса, события или явления, значение или поведение которого подлежит измерению или контролю телеметрической системой. В зависимости от скорости

изменения во времени ТПП делятся на медленно меняющиеся (ММТПП) и быстро меняющиеся (БМТПП). Первые характеризуются шириной спектра от 0 до 50 Гц, а вторые имеют верхнюю границу спектра от единиц до десятков и сотен килогерц [1]. В статье рассматриваются только параметры, принадлежащие к классу ММТПП.

Одной из основных задач в обработке телеметрии является задача выявления отклонений в поведении ТПП. Наиболее часто используемый подход, основанный на допусковом контроле значений, является низкоэффективным [1] в связи с тем, что параметры содержат большое количество

шумов и выбросов, значения которых выходят за границы допускового интервала. Для объектов, которые в рамках разных опытно-испытательных работ (ОИР) имеют схожее поведение, в качестве альтернативного может быть использован подход, основанный на сравнении параметров с набором эталонных параметров. Решение задачи сравнения временных рядов предполагает решение двух подзадач: определение способа формализованного описания временных рядов и метода их сравнения. Широко применяемые в настоящее время высокоуровневые представления временных рядов [2]–[8] и методы работы с ними не обеспечивают решение задачи сравнения телеметрических параметров. Основная проблема, возникающая при обработке телеметрических параметров, заключается в том, что параметры являются сложными нестационарными временными рядами, содержащими резкие переходы, соответствующие изменению состояния телеметрируемого объекта. Кроме того, в рамках разных ОИР время нахождения объекта в одном и том же состоянии может варьироваться, а значения одного и того же параметра могут передаваться с различным коэффициентом усиления исходного сигнала.

В статье предлагается метод сравнения ММТМП, основанный на использовании символьного представления временных рядов, применение которого позволяет эффективно решить проблемы, возникающие при обработке параметров.

Постановка задачи. Телеметрические параметры представляют собой временной ряд вида

$$C = c_1, \dots, c_n,$$

где c_i – измерение параметра в момент времени t_i ; n_i – количество измерений параметра. С информационной точки зрения наибольшую ценность для ММТМП представляют амплитудные составляющие параметра, а именно абсолютная величина, скорость, направление и форма изменения параметра.

ММТМП обладают кусочно-постоянной моделью поведения [1], которая предполагает описание временного ряда в виде последовательности участков, каждый из которых обладает набором постоянных характеристик:

$$M = \langle M_1(\alpha_1, \beta_1, \dots, \omega_1), \dots, M_k(\alpha_k, \beta_k, \dots, \omega_k) \rangle,$$

где $M_i(\alpha_i, \beta_i, \dots, \omega_i)$ – описание поведения параметра на i -м участке; $\alpha_i, \beta_i, \dots, \omega_i$ – набор характеристик, описывающих i -й участок. Формируемое описание временного ряда должно удовлетво-

рять условию: $\sum_{i=1}^k |M_i| = n$, где $|M_i|$ – длина i -го участка.

Для выявления отклонений в поведении ТМП оно сравнивается с эталонным поведением Σ , определенным для каждого параметра. Эталонное поведение задается с помощью одного или нескольких временных рядов:

$$\Sigma = S_1, \dots, S_m,$$

где $S_i = s_1^i, \dots, s_n^i$ – временной ряд, определяющий один из возможных вариантов поведения (шаблон поведения) параметра, соответствующего штатному состоянию телеметрируемого объекта; m – общее количество шаблонов.

Необходимо разработать метод сравнения ММТМП, который обеспечит:

а) возможность формирования высокоуровневого представления параметров \hat{C} на основе исходного временного ряда:

$$\Phi: C \rightarrow \hat{C}, \hat{C} = \hat{c}_1, \dots, \hat{c}_w, n \gg w,$$

где Φ – оператор, выполняющий перевод C в \hat{C} ; \hat{c}_i – i -й элемент высокоуровневого представления; w – общее количество элементов высокоуровневого представления;

б) возможность расчета расстояния d между парами параметров, для которых сформированы высокоуровневые представления: $\hat{C}: d(\hat{C}_i, \hat{C}_j)$,

где \hat{C}_i и \hat{C}_j – описания i -го и j -го параметров соответственно.

Описание метода символьного представления ТМП. В основу метода сравнения ММТМП положен метод SAX (Symbolic Aggregate approXimation), предложенный Е. Кеогом [9]–[11]. Метод SAX предполагает формирование символьного представления исходного временного ряда через промежуточное кусочно-постоянное представление (Piecewise Aggregate Approximation, PAA). Для построения PAA-представления данные разбиваются на W участков (сегментов) одинаковой длины. При формировании описания временного ряда рассчитывается среднее значение для каждого из сегментов. Применение PAA-представления в алгоритме SAX позволяет сократить размерность обрабатываемых данных и обеспечивает наличие нижней границы вычисляемых значений расстояний между временными рядами [12], [13].

Алгоритм 1. Построение РАА-представления.

Вход: $C = c_1, \dots, c_n$ – исходный временной ряд, w – количество. Выход: $\bar{C} = \bar{c}_1, \dots, \bar{c}_n$.

1 $l \leftarrow \text{floor}\left(\frac{n}{w}\right)$

2 для $i=1$ до w

3 $\bar{c} = \frac{1}{l} \left(\sum_{j=l(i-1)+1}^{li} c_j \right)$

4 конец для

При построении символьного представления на основе РАА-представления делается предположение о том, что нормированные значения исходного временного ряда имеют распределение, близкое к нормальному. Для сопоставления каждому из значений РАА-представления символа диапазон значений разбивается на равные области (уровни значений): $B = \beta_1, \dots, \beta_{\alpha-1}$, где B – множество уровней разбиения; β_i – граница между $(i-1)$ -м и i -м уровнями; $[\beta_{i-1}, \dots, \beta_i)$ – i -й уровень разбиения. Все уровни разбиения имеют одинаковую площадь, равную $1/\alpha$. Кроме того, предполагается, что $\beta_0 = -\infty$, $\beta_\alpha = +\infty$. Каждому уровню разбиения сопоставляется уникальный символ из конечного алфавита. Построенное разбиение обеспечивает возможность формирования однозначного символьного описания временного ряда.

При построении символьного представления телеметрических параметров с использованием алгоритма SAX необходимо учесть следующие особенности:

а) распределение исходных значений ММТМП не подчиняется нормальному закону;

б) символьные описания временных рядов при использовании различных коэффициентов усиления будут существенно различаться;

в) разбиение на равные интервалы при формировании промежуточного описания ММТМП может привести в тому, что в один сегмент будут включены измерения, соответствующие различным состояниям объекта и переходу между состояниями.

Учет перечисленных особенностей работы с телеметрическими параметрами обеспечивается за счет внесения следующих модификаций в алгоритм SAX:

а) осуществляется предварительная нормализация параметров таким образом, чтобы значения нормализованного временного ряда находились в интервале $[-1, 1]$; при нормализации временного

ряда в качестве минимального и максимального значений рассматривается медиана, вычисленная среди K максимальных и минимальных значений соответственно, что позволяет существенно снизить влияние шумов и выбросов на работу алгоритма;

б) в качестве граничных значений сегментов рассматриваются моменты, соответствующие сменам состояния телеметрируемого объекта.

Далее представлен модифицированный алгоритм формирования символьного представления \hat{C} для исходного временного ряда C .

Алгоритм 2. Построение символьного представления.

Вход: $C = c_1, \dots, c_n$ – исходный временной ряд; $P = p_1, \dots, p_w$ – w моментов смены состояния телеметрируемого объекта; $A = \alpha_1, \dots, \alpha_{k-1}$ – алфавит символьного представления длиной k ; $B = \beta_1, \dots, \beta_{k-1}$ – значения уровней. Выход: $\hat{C} = \hat{c}_1, \dots, \hat{c}_n$ – символьное представление.

1 $\hat{C} \leftarrow$ выполнить алгоритм 1 для C и P

2 $\bar{C}' \leftarrow$ нормализовать \bar{C}

3 для $i=1$ до w

4 $\hat{c}_i = a_j \Leftrightarrow \beta_{j-1} \leq \bar{c}'_i < \beta_j$

5 конец для

Результат конкатенации символов $\hat{c}_1, \dots, \hat{c}_w$ определяет символьное представление временного ряда \hat{C} .

Редакционное расстояние и алгоритм его построения. Расстояние Левенштейна (также редакционное расстояние или дистанция редактирования) между двумя строками определяется как минимальное количество операций вставки / удаления одного символа или замены одного символа на другой, необходимых для преобразования одной строки в другую [14]. Под редакционным предписанием понимается последовательность действий, необходимых для преобразования первой строки во вторую с использованием минимального количества операций. При вычислении редакционного расстояния учитываются цены операций, которые зависят от вида операции (вставка, удаление, замена) и/или от участвующих в ней символов [15].

Для вычисления редакционного расстояния предлагается использовать алгоритм Вагнера–Фишера [16]. В алгоритме приняты следующие обозначения: $\omega(a, b)$ – цена замены символа a на символ b ; $\omega(e, b)$ – цена вставки символа b ;

$\omega(a, e)$ – цена удаления символа a . В качестве метрики расчета расстояния между символами рассматривается разность значений границ уровней соответствующих символов. Тогда функция расчета цены замены символа a на символ b имеет вид

$$\omega(a, b) = \begin{cases} 0, & \text{если } a = b, \\ |\beta_a - \beta_b|, & \text{если } a \neq b, \end{cases}$$

где β_a и β_b – границы уровней значений.

Алгоритм 3. Алгоритм Вагнера–Фишера.

Вход: \hat{C}_1 , \hat{C}_2 – две сравниваемые строки, ω – функция, определяющая цену операций. Выход: D – редакционное расстояние между символическими представлениями \hat{C}_1 , \hat{C}_2 временных рядов C_1 и C_2 соответственно.

- 1 $D(0, 0) = 0$
- 2 для $j = 1$ до $|\hat{C}_2|$
- 3 $D(0, j) = D(0, j-1) + \omega(e, \hat{C}_2[j])$
- 4 конец для
- 5 для $i = 1$ до $|\hat{C}_1|$
- 6 $D(i, 0) = D(i-1, 0) + \omega(\hat{C}_1[i], e)$
- 7 для $j = 1$ до $|\hat{C}_2|$
- 8 $D(i, j) = \min \begin{cases} D(i-1, j) + \omega(\hat{C}_1[i], e), \\ D(i, j-1) + \omega(e, \hat{C}_2[j]), \\ D(i-1, j-1) + \omega(\hat{C}_1[i], \hat{C}_2[j]). \end{cases}$
- 9 конец для
- 10 конец для
- 11 вернуть $D(|\hat{C}_1|, |\hat{C}_2|)$

Оценки метода нечеткого сравнения телеметрических параметров. В методе построения символического представления ТМП для формирования РАА-представления необходимо просмотреть исходный временной ряд длиной n и вычислить среднее значение для ω сегментов, что тре-

бует временных затрат $O(n)$ и затрат по памяти $O(\omega)$. Перевод из РАА-представления в символическое представление осуществляется за линейное время и зависит от количества сегментов ω . Таким образом, суммарные временные затраты на формирование символического представления ТМП из исходного временного ряда составят $O(n + \omega)$. В связи с тем, что $n \gg \omega$, временные затраты, необходимые на выполнение метода, составляют $O(n)$.

Алгоритм Вагнера–Фишера, применяемый для сравнения временных рядов, в ходе работы осуществляет заполнение матрицы размером $|\hat{C}_1| \times |\hat{C}_2|$, что определяет время работы алгоритма и объем необходимой памяти, которые составляют $O(|\hat{C}_1| \cdot |\hat{C}_2|)$. Объем требуемой памяти может быть сокращен до линейного, если не сохранять промежуточные результаты работы алгоритма.

Разработанный метод приближенного сравнения телеметрических параметров на основе символического представления позволяет оценивать сходство анализируемых телеметрических параметров с ранее полученными параметрами от однотипных объектов, полученных при аналогичных условиях проведения ОИР. Реализация метода в системах интеллектуального анализа телеметрической информации позволила оперативно решать задачи оценки результатов проведенных ОИР. При обнаружении отклонений в поведении телеметрических параметров системой предоставляется информация о составе параметров, в которых выявлены отклонения, и моментах возникновения отклонений. Дальнейшее направление развития предложенного метода предполагает анализ возможности увеличения скорости работы метода за счет использования альтернативных алгоритмов вычисления редакционного расстояния между строками.

СПИСОК ЛИТЕРАТУРЫ

1. Назаров А. В. Современная телеметрия в теории и на практике: Учеб. курс. СПб.: Наука и техника, 2007.
2. Chan K., Fu A.W. Efficient Time Series Matching by Wavelets // In proc. of the 15th IEEE Int'l Conf. on Data Engineering, Sydney, Australia, Mar. 23–26, 1999. P. 126–133.
3. Faloutsos C., Ranganathan M., Manolopoulos Y. Fast Subsequence Matching in Time-Series Databases // In proc. of the ACM SIGMOD Int'l Conf. on Management of Data, Minneapolis, May 24–27, 1994. P. 419–429.
4. Geurts P. Pattern Extraction for Time Series Classification // In proc. of the 5th European Conf. on Principles of Data Mining and Knowledge Discovery, Freiburg, Germany, Sep. 3–7, 2001. P. 115–127.
5. Keogh E., Pazzani M. An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback // In proc. of the 4th Int'l Conf. on Knowledge Discovery and Data Mining, N.Y., Aug 27–31, 1998. P. 239–241.

6. Scargle J., Jackson B., Norris J. Adaptive Piecewise-constant Modeling of Signals in Multidimensional Spaces // In proc. of the Statistical Problems in Particle Physics, Astrophysics and Cosmology, SLAC, Stanford, California, Sep. 8-11, 2003. P. 157–162.
7. Kawahara Y, Sugiyama M. Change-Point Detection in Time-Series Data by Direct Density-Ratio Estimation // In Proc. of the SIAM Intern. Conf. on Data Mining, SDM, April 30-May 2, 2009. Sparks.
8. Turner R., Saatei Y, Rasmussen C. E. Adaptive Sequential Bayesian Change Point Detection // Temporal Segmentation Workshop at NIPS. 2009. Dec. P. 1–4.
9. A Symbolic Representation of Time Series, with Implications for Streaming Algorithms / J. Lin, E. Keogh, S. Lonardi, B. Chiu // In proc. of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, California. 13th June, 2003. P. 2-11.
10. Experiencing SAX: a Novel Symbolic Representation of Time Series / J. Lin, E. Keogh, S. Lonardi, B. Chiu // Data Mining and knowledge discovery. 2007. Vol. 15, № 2. P. 107–144.
11. Lin J., Keogh E., Fu. A.W. HOT SAX: Efficiently Finding the Most Unusual Time Series Subsequence // In proc. of the 5th IEEE Intern. Conf. on Data Mining, Houston, TX, Nov. 27-30, 2005. P. 226–233.
12. Keogh E., Chakrabarti K., Pazzani M. Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases // In proc. of ACM SIGMOD Conf. on Management of Data, Santa Barbara, May 21-24, 2001. P. 151–162.
13. Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases / E. Keogh, K. Chakrabarti, M. Pazzani, S. Mehrotra // J. of Knowledge and Information Systems. 2001. Vol. 3. P. 263–286.
14. Левенштейн В. И. Двоичные коды с исправлением выпадений, вставок и замещений символов // Докл. АН СССР. 1965. С. 845–848.
15. Гасфилд Д. Строки, деревья и последовательности в алгоритмах: Информатика и вычислительная биология / Пер. с англ. И. В. Романовского. СПб.: Невский Диалект; БХВ-Петербург, 2003.
16. Wagner R. A., Fischer M. J. The string-to-string correction problem // J. of the ACM. 1974. Vol. 21, № 1. P. 168–173.

N. A. Zhukova, I. S. Sokolov, A. V. Ekalo

THE METHOD OF FUZZY COMPARISON OF TELEMETRIC PARAMETERS BASED ON DATA MINING

This paper offers the method of fuzzy comparison of slowly changing telemetric parameters based on symbolical representation with use of the weighed edit distance. Application of edit distance allows to compare parameters displaced from each other by time and/or values, and to eliminate insignificant deviations in parameters behavior, caused by influence of external factors.

Data Mining, symbolical representation, telemetric parameters

004.42.032.24

М. А. Фирсов, С. А. Ивановский

Параллельная реализация алгоритма построения пересечения простых полигонов с использованием технологии CUDA

Описана параллельная реализация на CUDA модифицированного расширенного алгоритма Грейнера-Хорманна построения пересечения простых плоских полигонов. Приведены результаты экспериментального сравнения времён выполнения последовательной версии и версии на CUDA.

Вычислительная геометрия, пересечение простых полигонов, технология CUDA

Операция построения оверлеев (объединения, пересечения или разности) плоских полигонов является одной из основных в системах автоматизированного проектирования (САПР), ГИС в других графических системах. На основе построения оверлеев решается целый ряд других смежных задач [1].

Разработано множество последовательных алгоритмов пересечения полигонов [2], но пока мало адаптаций этих алгоритмов для выполнения на графических ускорителях. В данной статье описывается параллельная версия для CUDA модифи-