Expert Systems with Applications 42 (2015) 8078–8085

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A dynamic time warped clustering technique for discrete event simulation-based system analysis

Michael Johnstone *, Vu Thanh Le, James Zhang, Bruce Gunn, Saeid Nahavandi, Doug Creighton

Centre for Intelligent Systems Research, Deakin University, Waurn Ponds, Australia

ARTICLE INFO

Article history:
Available online 30 June 2015

Keywords:
Discrete event simulation
Simulation trajectory analysis
Dynamic time warping
Clustering
K-medoids

ABSTRACT

This paper introduces a novel approach for discrete event simulation output analysis. The approach combines dynamic time warping and clustering to enable the identification of system behaviours contributing to overall system performance, by linking the clustering cases to specific causal events within the system. Simulation model event logs have been analysed to group entity flows based on the path taken and travel time through the system. The proposed approach is investigated for a discrete event simulation of an international airport baggage handling system. Results show that the method is able to automatically identify key factors that influence the overall dwell time of system entities, such as bags that fail primary screening. The novel analysis methodology provides insight into system performance, beyond that achievable through traditional analysis techniques. This technique also has potential application to agent-based modelling paradigms and also business event logs traditionally studied using process mining techniques.

1. Introduction

Discrete event simulation (DES) is a powerful modelling methodology employed for the design and evaluation of complex and dynamic systems (Robinson, 2005) such as those found in manufacturing, logistics, health and communications (Ashour & Okudan Kremer, 2013; Qu & Meng, 2012; Yoo, Cho, & Yücesan, 2010).

The DES methodology abstracts real systems to a sequence of processes that act on entities flowing from process to process. The variability inherent in the real system is important for understanding system behaviour and included in the DES model. Analysis of in-system time profiles of entities within a system can be used to correctly gauge the impacts of variability on system performance. However as systems and their corresponding models become more complex, it becomes increasingly difficult to draw conclusions about system behaviour through manual review of in-system time profiles.

In this article we address the issue of complex DES output analysis with a generic, novel method that is simulation software agnostic. This approach complements existing simulation output data analysis techniques that are frequently built into simulation software. The proposed method identifies the traits of entities that contribute to overall system performance by combining dynamic time warping (DTW) and k-medoids clustering.

Typical methods for DES output analysis focus on process level metrics, with common examples including utilisation rates, waiting times and queue lengths (Ashour & Okudan Kremer, 2013; Fassi, Awasthi, & Viviani, 2012; Johnstone, Le, Nahavandi, & Creighton, 2009). Commercial DES software incorporates inbuilt analysers for these metrics (Rohrer & McGregor, 2002), as well as support for methods such as sensitivity analysis and design of experiments.

Alternate methods of analysis focus on the output of the simulation model to understand the dynamic behaviour exhibited by resources within the model. Kemper and Tepper (2009) employ simulation traces to provide insight into repetitive versus progressive behaviour for debugging the simulation model. Vasyutynskyy, Gellrich, Kabitzsch, and Wustmann (2010) similarly used simulation model event logs, combining these with a rule set, to evaluate key system indicators and identify areas for improvement.

These common analysis techniques operate at the local process level and provide information relating to an individual process or a particular local segment of the entire system. Often, a more global view of the system is desired when studying systems based on a flow of entities (Le, Zhang, Johnstone, Nahavandi, & Creighton,

* Corresponding author. Tel.: +61 3 5227 1104.
E-mail addresses: Michael.Johnstone@deakin.edu.au (M. Johnstone), Vu.Le@deakin.edu.au (V.T. Le), James.Zhang@deakin.edu.au (J. Zhang), Bruce.Gunn@deakin.edu.au (B. Gunn), Saeid.Nahavandi@deakin.edu.au (S. Nahavandi), DougC@deakin.edu.au (D. Creighton).

http://dx.doi.org/10.1016/j.eswa.2015.06.040
0957-4174/

2012; Zhang, Le, Johnstone, Nahavandi, & Creighton, 2012). Additional insight into overall system performance is needed and can be investigated by understanding how an entity's journey through the system impacts the overall system performance. New insights can be sought through system analysis from the perspective of the flow of entities as a time series.

By considering the travelled path of an entity through the simulated system, a DES model can generate a time series per entity. Dynamic time warping (Sakoe & Chiba, 1978) can then provide a means to measure the similarity between any two time series. This idea has previously been employed to discover business key performance indicators from data, when combined with Granger causality to determine if one indicator is a predictor of another (Peng, Sun, Rose, & Li, 2007). Peng et al. applied the method to a case study to find bottlenecks in the system and increase the throughput at that resource to improve system performance. Again, the focus was on processes within the system rather than analysing the flow of items through the system. Kim, Lee, Lim, Seo, and Kang (2014) applied stepwise lower-bounded dynamic time warping to the problem of handwriting recognition from gyroscope-generated data. The performance was shown to be superior to the conventional methods such as hidden Markov model (HMM). Górecki and Łuczak (2015) combined two distances in their approach: the DTW distance between multivariate time series and the distance between derivatives of the time series. The results on 18 datasets demonstrated the effectiveness of the proposed approach. It should be noted that the approach in Górecki and Łuczak (2015) can only be applied to a continuous signal, as derivatives are involved. Using DTW as a distance function clustering technique, can be applied to reveal and visualise the internal structure of the collected data and hence the system that generated the data. Izakian, Pedrycz, and Jamal (2015) proposed three alternatives to fuzzy clustering of time series data, based on the DTW distance. The first is the DTW-based averaging technique applied to the fuzzy C-means clustering. The second approach used fuzzy C-medoid clustering and the third alternative is a hybrid technique of the former two methods.

The dynamic time warping technique has been used in a wide and diverse range of applications concerning time series analysis, such as health and rehabilitation (Su, Chiang, & Huang, 2014), robot trajectory learning (Vakanski, Mantegh, Irish, & Janabi-Sharifi, 2012) and econometrics (Tsinaslanidis & Kugiumtzis, 2014). However an important and potential fruitful application domain, the analysis of discrete event simulation output, especially in the area of logistics and manufacturing, where a customer goes through various stages of a complex system to receive service, has not received much attention. In this paper, we propose to combine DTW with k-medoids clustering, with the aim to understand how an entity's journey through the system contributes to features of the overall performance of the system. This is achieved by developing causal reasons for the clustering which has been observed to occur. This leads to insights on system behaviour and lends itself to potential approaches to improve overall system performance. Clustering has been successfully used to analyse large log files with the aim of summarising the content of the log files to avoid overwhelming the analyst (Vaarandi, 2003). We use unsupervised clustering methods to tease out system features in order to gain insight into system behaviour. In short, the DTW and clustering techniques provide summarised information that helps explain system behaviour, and identify areas for system improvement.

The remainder of this paper is structured as follows. First, the approach to analysing DES output event logs using DTW is described, followed by an overview of the k-medoids clustering technique as it is applied to the output from the DTW technique. Next we apply the two methods to a case study simulated system and evaluate the results. We conclude with a discussion on the merits of our approach and indicate areas for further study.

2. Dynamic time warping for simulation output analysis

An understanding of how the effects of processing entities, as they flow through the system, contributes to overall system behaviour is desired. Hence, there are spatial considerations of how an entity has traversed the system and temporal considerations of what time events occurred to change the state of the entity in the system.

DTW has proven successful in determining the similarity of two time series and representing this as a distance measure in a variety of fields, including speech recognition, data mining and manufacturing (Keogh & Pazzani, 2000; Myers, Rabiner, & Rosenberg, 1980; Young-Seon, Seong-Jun, & Jeong, 2008). Fast DTW (Salvador & Chan, 2007), an implementation of DTW, is applied to the DES model output, which is a series of events and corresponding timestamp for each entity. Events relate to the location and processing status of entities as they flow through the system. Hence, location information and the results of processing are encoded in the simulation output and is aimed at providing insight into both the path an entity may take and also the results of processing. DTW is used to consider different paths through the system, processing times through the system and the alternative processing incurred by entities as they traverse the system.

At the conclusion of a simulation run, the event path for all entities is output into a single file, which can be parsed into a format suitable for fast DTW. A matrix is then constructed to hold the similarity index between each entity pair in preparation for the clustering stage.

The formalisation to construct the time series from DES output follows Salvador and Chan (2007):

Let N be the number of entities that complete processing and output a path trace.

Let E be the set of time series $E_1, E_2, \ldots E_N$.

A time series consists of a sequence of time t and event e pairs:

$$E_n = (t_1, e_1), (t_2, e_2), \ldots, (t_{M_n}, e_{M_n})$$

where M_n is the number of events in the path of E_n.

For two entities, E_x and E_y

$$E_x = (t_i, e_i) \quad i = 1, 2, \ldots, M_x \quad (x \in N)$$

$$E_y = (t_j, e_j) \quad j = 1, 2, \ldots, M_y \quad (y \in N)$$

the warp path W is constructed with length K

$$W = w_1, w_2, \ldots, w_K$$

$$\max(M_x, M_y) \leqslant K < M_x + M_y$$

and is indexed via $w_k = (i, j)$, where i and j are indexes from the time series E_x and E_y.

The distance between the two time series is defined by the shortest warp path distance given by:

$$Dist(W) = \sum_{k=1}^{K} Dist(w_{ki}, w_{kj})$$

Typically the Euclidean distance is used between the two data points (Salvador & Chan, 2007). We propose a variation, dependent on the event stamp:

$$Dist_k = \begin{cases} |t_{ki} - t_{kj}|, & \text{if} \quad e_{ki} = e_{ji} \\ 1 + |t_{ki} - t_{kj}|, & \text{if} \quad e_{ki} \neq e_{ji} \end{cases}$$

If the event types match, the distance becomes the Manhattan distance between the event times, else the distance is one plus the