Discovering Causality in Event Time-Series

Pavel Loskot
ZJU-UIUC Institute
Haining, China
e-mail: *pavelloskot@intl.zju.edu.cn*

***Abstract*—The event time-series can accurately describe the behavior of many dynamic systems. The challenge is that the events are categorical variables, so they cannot be analyzed by the existing statistical methods developed for numerical time-series. In order to infer the causally related events, in this paper, it is proposed to assume the empirical conditional probabilities of nearly certain and nearly uncertain events. Moreover, since the event ordering is usually locally irrelevant, the event sequences can be transformed into the event sets or multi-sets with appropriately defined distance metrics. The event sequences having a zero distance can be then assumed to be causally equivalent. The distance metrics are also used in matrix profile analysis of event time-series. Numerical examples are studied for chemical reaction events generated in stochastic simulations of biochemical molecular systems. Even though the proposed framework for discovering the causally related event sequences can be readily fully automated, they still need to be properly interpreted in the context of relevant domain knowledge.**

***Keywords*—causal; dynamic system; event; matrix profile; state-space; time-series**

I. Introduction

Traditional signal processing and machine learning mainly exploit statistical associations within data. However, it is well known that a strong association is neither necessary nor sufficient for causality, for example, due to confounding. At the same time, a weak association cannot rule out the causality. Recently, there has been a great interest in developing data models and processing methods, which are interpretable [1].

The cause and effect are central to scientific hypotheses testing, experiment design, and to generate the prescriptive analytics of engineering systems. It is possible to only consider whether the cause-effect exists without determining its direction or strength. The causal relationships can be represented as Structural Causal Models (SCM) [2]. The SCM can be created from a prior knowledge, or inferred by performing statistical independence tests on the data. An important question is whether the SCM can be determined from the observed data, and whether such a SCM is unique. The SCM can be converted into a Bayesian network using do-calculus [2].

Different approaches were adopted in the literature to obtain causal models of time-series data [3]–[11]. Granger causality decides whether the past values of a time-series can improve the prediction of future values of another time-series. However, this type of causality cannot be used for time-series with instantaneous effects, or when sub-sampling of time-series may mask the causal relationships. The intervention causality enforces a change in the time-series value at a particular time instant, and then the change can be evaluated as an Average Causal Effect (ACE). In supervised and semi-supervised machine learning, the labels of data can be assumed to be a cause of data features (i.e., the effects). It enables to automatically label data as well as to repair incorrect labels. However, all these methods normally assume numerical data.

In [12], the causality is induced by changes in the interaction covariances. The methods for evaluating causal intervention of non-randomized, small-size treatments are surveyed in [13]. A state-space SCM for causal inference in time-series data was studied in [14]. A causal graph discovery over multiple related datasets was proposed in [15]. The limitations of convergent cross-mapping in performing the causal inference were investigated in [16]. The temporal trends in data need to be identified before performing the causal inference as shown in [17]. The causal analysis of small sample sizes was performed in [18] by studying state-space attractors of nonlinear dynamical systems. However, none of these works seem to have considered the causal inference for categorical data.

In this paper, our goal is to discover causal relationships within categorical time-series. Such series may represent the events occurring in control and monitoring of dynamic systems. The events cannot be often directly detected, but must be indirectly inferred except in computer simulations of dynamic systems. The events usually incur changes in the system internal states. It is extremely useful to understand what caused these changes, and to make more robust predictions about the anticipated future changes (effects). Moreover, the event time-series can be partitioned into shorter sequences. The task is then to determine the causality between the pairs of the event sequences. In addition, since the event ordering is locally irrelevant, it is proposed to transform the event sequences into the event sets or multi-sets.

More importantly, the cause-effect relationship is newly defined here assuming the conditional probability of nearly certain and nearly uncertain events. This probability is estimated empirically as a relative frequency of occurrence of particular event sequences. Even though such a notion of causality is incomplete, as many event sequences are conditionally neither certain nor uncertain, this approach has the advantage of its implementation simplicity, and it can be fully automated. Moreover, various distance measures [19]–[22] can be used to define equivalences among the event sequences, which can increase the number of these sequences classified as being causally related by our definition of causality. The distance metrics also enable the matrix profile analysis, a versatile framework used for the pattern discovery in time-series data.

 ISBN: 978-1-61208-970-6

Numerical examples are obtained for a biochemical reaction network, where the events represent chemical reactions. The history of chemical reactions are recorded by modifying the downloaded open-source simulation. The event time-series processing and visualization pipeline is implemented using a C++ code and the custom scripts in Python and Bash.

The rest of this paper is organized as follows. Section II describes a common model of dynamic systems with event-driven changes of observations and internal states. The proposed causal framework for analyzing the event time-series is introduced in Section III. Numerical examples are briefly presented in Section IV. The limitations of our work are discussed and the paper is concluded in Section V.

II. System Model

Consider a dynamic system described by transitions between the consecutive stationary states $\boldsymbol{z}_{t-1} \in \mathcal{Z}$ and $\boldsymbol{z}_t \in \mathcal{Z}$ due to periodically occurring events $e_t \in \mathcal{E}$, i.e.,

$$\boldsymbol{z}_t = e_t(\boldsymbol{z}_{t-1}, \boldsymbol{z}_{t-2}, \ldots)$$

where t denotes a discrete time index. The system observations are defined by a generally non-linear function,

$$\boldsymbol{y}_t = O(\boldsymbol{z}_t, \boldsymbol{z}_{t-1}, \ldots)$$

of the current and the previous system states including any intrinsic and extrinsic noises (the latter not shown explicitly). The values $\boldsymbol{y}_t$ may not be available for all indexes t due to practical measurement constraints; this corresponds to uniform or non-uniform sub-sampling of the observations $\boldsymbol{y}_t$.

In this paper, it is assumed that, (1) the system model is memoryless, i.e., $\boldsymbol{z}_{t+1} = e_t(\boldsymbol{z}_t)$, and, $\boldsymbol{y}_t = O(\boldsymbol{z}_t)$, and, (2) the observations are perfect, i.e., $\boldsymbol{y}_t = \boldsymbol{z}_t$, and available for all t. Such Markovian and noise-free observation assumptions greatly simplify our reasoning, and they are satisfied for the system studied in Section IV. Furthermore, the events e_t are represented as categorical variables, such that, $e_t \in \{0, 1, 2, \ldots\}$. Under these assumptions, the transitions between the observations $\boldsymbol{y}_t$ and the events $\boldsymbol{e}_t$ are depicted in Figure 1. In particular, Figure 1 indicates that different events affect different components in the observed vector, $\boldsymbol{y}_t$. However, the practical constraints on the observations $\boldsymbol{y}_t$, for example, to ensure that, $\boldsymbol{y}_t \geq 0$ for $\forall t$, may enforce a dependency (i.e., a memory) among the successive events e_t. Note also that both the events and the observations are normally dependent on a number of other parameters, which is not explicitly considered in our model description.

The memoryless assumption implies the following property of the event-based modeling of dynamic systems.

Theorem 1: Given the observation $\boldsymbol{y}_t$ at time t, the ordering of $(m+1) > 0$ events in the sequence, $(e_t, e_{t+1}, \ldots, e_{t+m})$, does not affect the observation $\boldsymbol{y}_{t+m}$ at time $(t+m)$.

Theorem 1 asserts that the same observation $\boldsymbol{y}_{t+m}$ is produced for any arbitrary ordering of a particular sequence of events. However, this does not guarantee that all the event orderings satisfy all the observation constraints; for instance, the natural constraints that the observations are always non-negative, or do not exceed a certain value, may be temporarily violated.

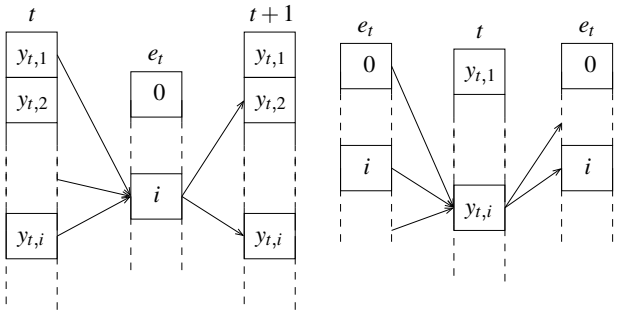

Figure 1. The changes in different components of observations $\boldsymbol{y}_t$ affected by different events $e_t \in \mathcal{E}$.

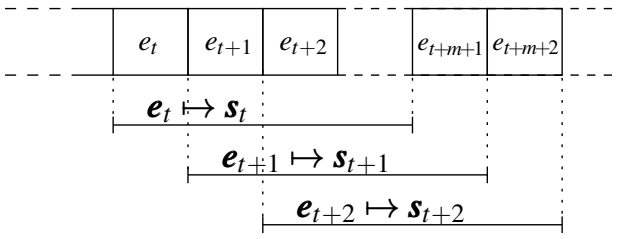

Figure 2. The sequences $\boldsymbol{e}_t$ of $(m+1)$ events mapped to (multi-) sets $\boldsymbol{s}_t$.

Consequently, the sequences of events, $\boldsymbol{e}_t = (e_t, e_{t+1}, \ldots, e_{t+m})$, can be assumed to be multi-sets (i.e., the same events can appear multiple times, but their ordering is irrelevant), or ordinary sets (the repeated elements are removed). The corresponding (multi-) sets are denoted as, $\boldsymbol{s}_t$, and they can be created by sliding-window partitioning of the original event time-series as shown in Figure 2.

III. Analysis of Event Time-Series

Recall that the events e_t are categorical variables, which can be mapped to non-negative integers $\mathcal{E}$. Since such a mapping is rather arbitrary, and assumed purely for a representation convenience, it cannot be used for evaluating statistical properties of the time-series, $\{e_t\}_t$. Assuming instead the sequence of (multi-) sets, $\boldsymbol{s}_t \in \mathcal{S}$, where $\mathcal{S} \subseteq \mathcal{E} \times \cdots \times \mathcal{E} = \mathcal{E}^{m+1}$, we can examine the probability mass function as well as define various distance measures involving $\boldsymbol{s}_t$. The former approach will be used to identify the causal relationships between pairs of event sequences. The latter approach enables a flexible matrix profile analysis of the event time-series.

A. Causality Between Event Sequences

Our objective is to determine a possible causal relationship between pairs of consecutive but non-overlapping event sequences. Thus, given $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, $j = i+m+1$, i.e., $\boldsymbol{e}_i \cap \boldsymbol{e}_j = \emptyset$ (empty set), decide, whether the event sequence $\boldsymbol{e}_i$ causes the event sequence $\boldsymbol{e}_j$ (causal learning), or whether the event sequence $\boldsymbol{e}_j$ is an effect of the event sequence $\boldsymbol{e}_i$ (anti-causal learning). One plausible and commonly used strategy is to construct a SCM, and fit it to the data (the event sequences).

 ISBN: 978-1-61208-970-6

The SCM analysis can be combined with interventions and do-calculus to find the causes and effects of specific event sequences. In this paper, we instead propose the following strategy to identify some, but not all pairs of causally related event sequences.

Definition 1: The event sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, $j > i$, have a cause-effect relationship, provided that their conditional probability,

$$\Pr(\boldsymbol{e}_j|\boldsymbol{e}_i) \to 1. \tag{1}$$

In such a case, the event sequence $\boldsymbol{e}_i$ is said to be a cause of $\boldsymbol{e}_j$. Equivalently, the event sequence $\boldsymbol{e}_j$ is an effect of $\boldsymbol{e}_i$.

The conditional probability (1) represents the likelihood of the event sequence $\boldsymbol{e}_i$, given the event sequence $\boldsymbol{e}_j$. Thus, the prior distribution of $\boldsymbol{e}_i$ is ignored in Definition 1, and so is the joint distribution $\Pr(\boldsymbol{e}_i, \boldsymbol{e}_j)$. This can be justified by noting that a cause-effect relationship is generally asymmetric. Even though there are also cases where $\boldsymbol{e}_i$ causes $\boldsymbol{e}_j$, and at the same time, $\boldsymbol{e}_j$ causes $\boldsymbol{e}_i$, in general, $\Pr(\boldsymbol{e}_j|\boldsymbol{e}_i) \neq \Pr(\boldsymbol{e}_i|\boldsymbol{e}_j)$.

The reason for assuming the conditional probability in Definition 1 to be nearly but not exactly equal to 1 is that the equality is too restrictive and almost never achievable in practice. Moreover, the conditional probability $\Pr(\boldsymbol{e}_j|\boldsymbol{e}_i)$ close to 1 indicates that, given $\boldsymbol{e}_i$, there are only a few possible event sequences $\boldsymbol{e}_j$ following $\boldsymbol{e}_i$ (i.e., the number of such sequences $\boldsymbol{e}_j$ is significantly smaller than the number of all possible event sequences observed).

More importantly, the size of the event space, $\mathcal{E}$, expressed as the cardinality, $|\mathcal{E}|^{(m+1)}$, is often much smaller than the volume of the corresponding observations $\boldsymbol{y}$ in a $(m+1)$-dimensional Euclidean or some other space. This is the key reason for analyzing the sequences of events for inferring the causality rather than directly processing the sequences of observations. Although in most practical scenarios, it is the observations that are available, whereas the internal events can only be inferred from these observations, which is prone to decision errors. However, computer simulations are a notable exception, and they will be utilized in Section IV.

The conditional probability of the event sequences in Definition 1 close to 1 is only one significant case, which can be readily causally interpreted. The other such significant case is represented by the conditional probability being close to 0. This leads to the following definition of causality between two event sequences.

Definition 2: The event sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, $j > i$, have no cause-effect relationship, provided that their conditional probability,

$$\Pr(\boldsymbol{e}_j|\boldsymbol{e}_i) \to 0 \quad \text{and} \quad \Pr(\boldsymbol{e}_i|\boldsymbol{e}_j) \to 0. \tag{2}$$

The sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are then said to be causally unrelated.

Assuming both conditional probabilities in Definition 2 is necessary in order to ensure that neither $\boldsymbol{e}_i$ nor $\boldsymbol{e}_j$ can be a cause of the other. Moreover, unlike Definition 1, it is much more likely to find the pairs of event sequences having very small or even exactly zero conditional probabilities.

The conditional probabilities of event sequences may sometimes be available from the analysis of a Bayesian model derived from the SCM. However, in many practical scenarios, these probabilities must be empirically estimated from the event time-series data. In such a case, the event sequences $\boldsymbol{e}_t$ of $N = (m+1)$ events are first created by a sliding-window partitioning of the original event time-series. In order to enable Definitions 1 and 2 of causality, the sequences $\boldsymbol{e}_t$ are further subdivided into two disjoint sub-sequences (omitting the time index for brevity), $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ of N_1 and N_2 events, respectively, so that,

$$\boldsymbol{e}_t = \boldsymbol{e}_i \cup \boldsymbol{e}_j, \quad \boldsymbol{e}_i \cap \boldsymbol{e}_j = \emptyset$$

and $N = N_1 + N_2$, $N_1 = |\boldsymbol{e}_i|$, $N_2 = |\boldsymbol{e}_j|$, and importantly, all the events in $\boldsymbol{e}_i$ precede the events in $\boldsymbol{e}_j$. The corresponding (multi-) sets are denoted as $\boldsymbol{s}_i$ and $\boldsymbol{s}_j$, and they are referred to as the left and the right event (multi-) sets, respectively.

Define a 2D counter (matrix), $C_{i,j}$, of the number of the unique left and right sub-sequences, $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, composing the event sequences, $\boldsymbol{e}_t$. The conditional probabilities (1) and (2) can be then estimated as,

$$\Pr(\boldsymbol{e}_j|\boldsymbol{e}_i) \approx C_{i,j}/K_i \tag{3}$$

where K_i denotes the number of times a specific event subsequence $\boldsymbol{e}_i$ was observed, i.e., $K_i = \sum_j C_{i,j}$.

There are, however, two issues with the causality in Definitions 1 and 2. The first problem is that the number of subsequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ satisfying (2) and especially (1) can be rather small in comparison to the total number of all observed event sub-sequences. This leaves out most other pairs of event sub-sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$, for which their causal relationship cannot be determined using Definitions 1 and 2, since their conditional probability, $0 < \Pr(\boldsymbol{e}_j|\boldsymbol{e}_i) < 1$. The second problem is that identifying the rarely occurring, causally related subsequences using the estimator (3) becomes less accurate, unless sufficiently long event time-series are available.

In order to overcome these issues, we can exploit the mapping of event sequences to event (multi-) sets, as discussed in Section II. It allows us to define various notions of distances between the event sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ as follows. Let d_0 be the Hamming distance between $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$. Then, any of the following expressions can be assumed as a distance metric between the event sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$.

$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d_0 - |\boldsymbol{s}_i \cup \boldsymbol{s}_j| \tag{4a}$$
$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d_0 - |\boldsymbol{s}_i \cap \boldsymbol{s}_j| \tag{4b}$$
$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d_0 - (|\boldsymbol{s}_i| + |\boldsymbol{s}_j|) \tag{4c}$$
$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d_0 - \max(|\boldsymbol{s}_i|, |\boldsymbol{s}_j|) \tag{4d}$$
$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = \max(|\boldsymbol{s}_i|, |\boldsymbol{s}_j|) - \min(|\boldsymbol{s}_i|, |\boldsymbol{s}_j|) \tag{4e}$$
$$d(\boldsymbol{e}_i, \boldsymbol{e}_j) = \min(|\boldsymbol{s}_i \setminus \boldsymbol{s}_j|, |\boldsymbol{s}_j \setminus \boldsymbol{s}_i|). \tag{4f}$$

Thus, always, $d(\boldsymbol{e}_i, \boldsymbol{e}_j) \geq 0$, $d(\boldsymbol{e}_i, \boldsymbol{e}_j) = d(\boldsymbol{e}_j, \boldsymbol{e}_i)$, and $|\boldsymbol{s}_i| \leq |\boldsymbol{e}_i|$.

Furthermore, in order to increase the number of occurrences of event sequences which are either causally related by Definition 1, or causally unrelated by Definition 2, we can assume

any of the distance metrics (4a)-(4f) to introduce the following notion of equivalent event sequences.

Definition 3: The event (sub-) sequences $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are said to be equivalent, provided that their distance, $d(\boldsymbol{e}_i, \boldsymbol{e}_j) = 0$.

It is clear that by discarding the event ordering and keeping only the unique events in event sets as assumed in the distance metrics (4a)-(4f), the number of equivalent event sub-sequences can grow substantially, and so does the number of event sub-sequence pairs satisfying either Definition 1 or Definition 2.

The effect of assuming the equivalent event sub-sequence is illustrated in Figure 3. In the first step, the unique event sub-sequence pairs $\boldsymbol{e}_i$ and $\boldsymbol{e}_j$ are identified, and their multiplicities, $C_{i,j}$, are counted. Using Definition 3, the equivalent event sub-sequences can be identified among the right or the left sub-sequences representing either the potential event causes, potential event effects, or both. The equivalent event sub-sequences are then merged (blue boxes in Figure 3), and the counters $C_{i,j}$ used in (3) are updated accordingly.

More specifically, let I_u, $u = 1, 2, \ldots$, be the sets of indices of the equivalent left event sub-sequences $\boldsymbol{e}_i$ representing possible causes, and $\mathcal{J}_v$, $v = 1, 2, \ldots$, are similar such index sets for the right event sub-sequences $\boldsymbol{e}_j$, representing the possible effects. The event sub-sequence counters in (3) are updated due to the left and the right merges as,

$$C'_{i,j} = \sum_{i \in I_u} C_{i,j}, \quad C'_{i,j} = \sum_{j \in \mathcal{J}_v} C_{i,j}.$$

Consequently, it then becomes much more likely that some pairs of the equivalent event sub-sequences have their conditional probability close to 1 (as estimated by their relative occurrences), so they can be assumed to be causally related by Definition 1. On the other hand, merging the equivalent event sub-sequences and aggregating the counters make it somewhat less likely that the condition of non-causal relationship in Definition 2 would be satisfied. These causal decisions are also greatly affected by a specific choice of the distance metric.

B. Matrix Profile Analysis of Event Time-Series

The canonical matrix profile effectively shows the minimum distances between constant length sequences, which are created by a sliding-window partitioning of the original time-series data. The distance calculations in the matrix profile are greatly optimized to allow processing of very long sequences of data. These calculations can be readily parallelized, for example, using a MapReduce algorithm. The matrix profile is mainly used to identify common patterns (motifs) as well as rare patterns (discords), and also to identify time instances when the distance-based sequence statistics have changed.

Even though the events are represented as categorical rather than numerical variables, the distance metrics (4a)-(4f) can be directly used in calculating the matrix profile of the event time-series. The choice of the actual distance metric strongly affects the resulting matrix profile, although less than one might expect. However, it is still useful to compare the matrix profiles for different values of the sequence lengths.

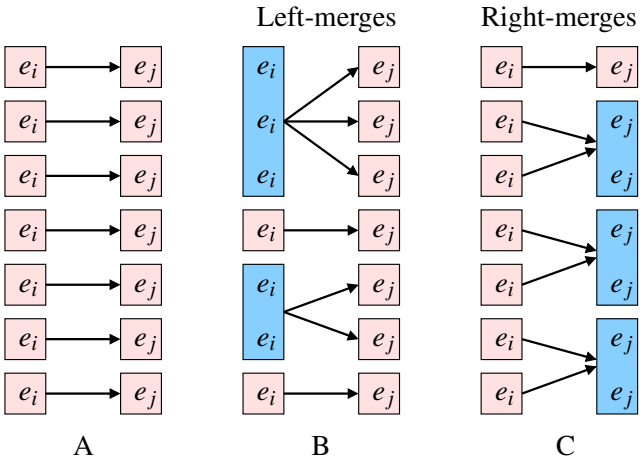

Figure 3. A: Original pairs of the unique event sub-sequences. B: Left-mergers of event sub-sequences as potential causes. C: Right-mergers of event sub-sequences as potential effects.

More importantly, the original matrix profile only displays the minimum distance values for each sliding-window sub-sequence. However, other statistics may also be useful. For example, the distribution of sub-sequence mutual distances provides a more global view, whereas the multiplicity of the smallest distances for each sub-sequence may be as informative in some applications as the actual smallest distance values.

IV. Case Study: Biochemical Reaction Networks

Biochemical reaction networks represent dynamic systems that undergo changes in copy counts (or, equivalently, concentrations) of chemical species due to chemical reaction events [23]. The number of chemical reactions is often much larger than the number of chemical species. The corresponding chemical kinetics can be stochastically described using a Chemical Master Equation (CME) [24]. The CME is usually solved by a Monte Carlo simulation [25], which tracks the time-evolution of the chemical species counts. More importantly, we assume a so-called well-stirred system, i.e., the spatial distribution and the diffusion of chemical molecules are ignored.

The models of chemical reaction systems may involve chemical species containing multiple binding sites [26]. Enumerating all chemical reactions for every binding site is impractical due to the combinatorial complexity of the resulting chemical reaction network. The network-free algorithms exploit the reaction (meta-) rules to effectively describe the groups of reactions without a need to enumerate all the reactions explicitly [27].

A. Numerical Experiments

Numerical experiments were obtained for an antigen receptor signaling regulating the activity and fate of the B-cells [28]. The corresponding model (referred to as BCR model) consists of 32 molecule types, 158 reaction rules, and 129 model parameters. The extracted full model contains 1,124 chemical species and 24,390 chemical reactions. The model was simulated in BioNetGen software [29], [30].

 ISBN: 978-1-61208-970-6

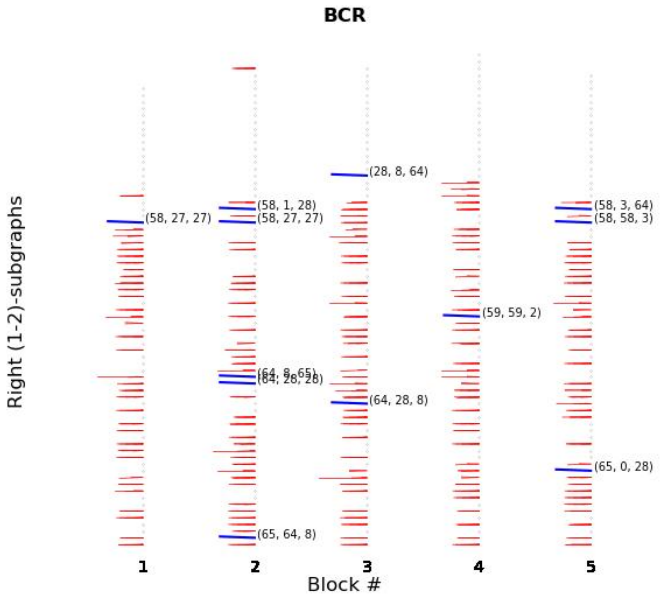

Figure 4. The estimated conditional probabilities $P_{j,i}$ of $N = (1+2)$ event sub-sequences. Black: $P_{j,i} < 0.1$, Red: $P_{j,i} < 0.9$, Blue: $P_{j,i} \geq 0.9$.

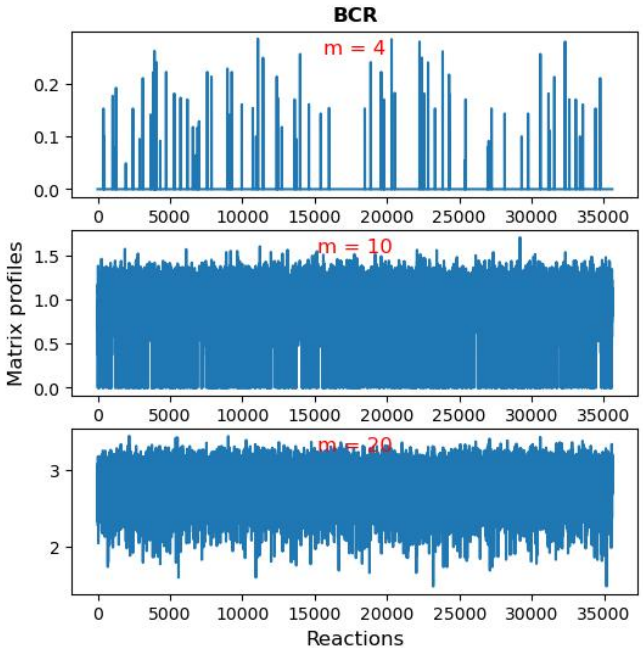

Figure 5. A canonical matrix profile assuming the minimum distances (4a), for sub-sequences of 4, 10 and 20 reaction events.

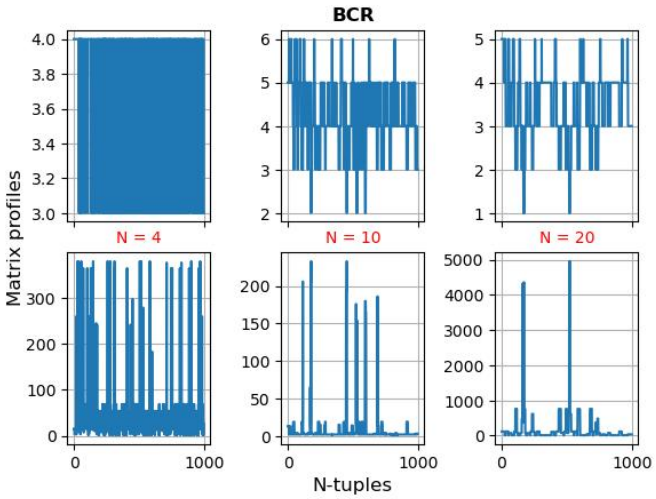

Figure 6. The matrix profile of the maximum distances (4f) (top-row) and their multiplicity (bottom-row) for sub-sequences of 4, 10 and 20 events.

BioNetGen is an open source software offering its own model description language to specify chemical reaction systems. The model description file is processed by a Perl script in order to generate the more complete system model given in System Biology Markup Language (SBML). The SBML file is then simulated in NFSim [31]; an open source software written in C++ [32]. We have modified NFSim to enable recording of the history of all reaction events in the course of the simulations. The trajectories of chemical species counts were simply discarded. The generated event time-series were processed, and visualized by the custom-made scripts written in Python. The overall process of performing the simulations, processing the event time-series, and generating the plots was fully automated mainly using the Bash scripts.

Simulating the BCR model over 100 simulation seconds resulted in 3,634,390 reaction events involving 35 reaction types. The reaction events can be naturally divided into 100 blocks over one second intervals. The sliding-window event sub-sequences were then formed and processed. The distinct frequencies of occurrence of the event N-tuples were observed, and they allow their clustering into multiple distinct classes.

Due to space limitations, only the following three plots are shown. Figure 4 visualizes the estimated reverse conditional probabilities, $\Pr(\boldsymbol{e}_i|\boldsymbol{e}_j)$ for the first five blocks (i.e., for the events $\boldsymbol{e}_i$ occurring before $\boldsymbol{e}_j$), assuming $|\boldsymbol{e}_j| = 2$ and $|\boldsymbol{e}_i| = 1$, i.e., $N = 3$. The reactions in each column in Figure 4 have the same ordering to indicate that some event patterns can be considered causal (according to our Definitions 1 and 2) in some blocks, but not in other blocks. Furthermore, in Figure 4, the right event sub-sequences were combined assuming the sub-sequence equivalences with the distance metric (4a). The line coloring is described in the caption of Figure 4.

Figure 5 depicts a canonical matrix profile assuming sub-sequences of 4, 10 and 20 events. The profile was calculated using the Python module stumpy [33]. It can be seen that the profile becomes more dense, and its mean moves away from zero with the increasing sub-sequence length. Finally, Figure 6 shows the matrix profile assuming the maximum (instead of minimum) values of the distance metric (4f) between any pair of sub-sequences of 4, 10 and 20 events, respectively (top-row), and the multiplicity of these maximum distance values (bottom-row). This illustrates how the choice of the distance metric greatly affects the shape of the matrix profile.

 ISBN: 978-1-61208-970-6

V. Discussion and Conclusion

The sliding-window event sequences were split into the left and the right event sub-sequences. The causality has been defined here as the pairs of nearly certain or nearly uncertain event sequences. The level of certainty can be evaluated empirically by measuring the corresponding conditional probabilities. Since ordering of events is locally irrelevant, it is useful to transform the event sub-sequences into event (multi-) sets, for which various distance metrics can be defined.

The distance metrics can be utilized to obtain the matrix profiles of event time-series. Our numerical experiments demonstrate that matrix profile is a rather general and flexible framework for analyzing numerical as well as categorical time-series, and conveniently visualizing their statistics.

Even though this paper focuses on analyzing the short sequences of consecutive events, the events neither have to be consecutive, nor short. However, assuming non-consecutive events make the pattern space to be combinatorially much larger, and the longer the event sequences, the less likely it is to identify those that can be considered to be statistically certain or uncertain. The causality analysis may also involve both the events and the observations. This can lead to explainable Monte Carlo simulations, provided that causally related (or unrelated) sub-sequences are identified and properly interpreted in a given domain [34], which can be challenging.

The simulation software adopted and the programming scripts produced to analyze state-space models of biochemical reaction networks allow fully automated processing of the recorded event time-series. It allows generating a large number of diverse plots for different models across different numerical experiments, although automated interpretation of the identified causal events may again be rather challenging.

Acknowledgment

This research was funded by a start-up research grant provided by Zhejiang University.

References

[1] J. Pearl, "The seven tools of causal inference, with reflections on machine learning," *Communications of the ACM*, vol. 62, no. 3, pp. 54–60, Mar. 2019.

[2] ——, "Causal inference in statistics: An overview," *Statistics Surveys*, vol. 3, pp. 96–146, 2009.

[3] K. Hlaváčková-Schindlera, M. Paluš, M. Vejmelka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, pp. 1–46, 2007.

[4] M. Eichler, *Causal Inference in Time Series Analysis*. Wiley, 2012, ch. In Causality (eds W.A. Shewhart et al.).

[5] A. Papana, C. Kyrtsou, D. Kugiumtzis, and C. Diks, "Simulation study of direct causality measures in multivariate time series," *Entropy*, vol. 15, pp. 2635–2661, 2013.

[6] D. Lopez-Paz, K. Muandet, B. Schölkopf, and I. Tolstikhin, "Towards a learning theory of cause-effect inference," in *ICML'15*, vol. 37, 2015, pp. 1452–1461.

[7] Y. Chikahara and A. Fujino, "Causal inference in time series via supervised learning," in *IJCAI'18*, 2018, pp. 2042–2048.

[8] R. M. et al., "Causal inference for time series analysis: problems, methods and evaluation," *Knowledge and Information Systems*, vol. 63, no. 12, pp. 3041–3085, Dec. 2021.

[9] Y. Hmamouche, A. Casali, and L. Lakhal, "A causality based feature selection approach for multivariate time series forecasting," in *DBKDA*, 2017, pp. 1–6.

[10] Y. Huang and S. Kleinberg, "Fast and accurate causal inference from time series data," in *Int. Florida AI Research Soc. Conf.*, 2015, pp. 1–6.

[11] J. R. at al., "Inferring causation from time series in Earth system sciences," *Nature Communications*, vol. 10, no. 2553, pp. 1–13, Jun. 2019.

[12] R. J. Prill, R. Vogel, G. A. Cecchi, G. Altan-Bonnet, and G. Stolovitzky, "Noise-driven causal inference in biomolecular networks," *PLoS ONE*, vol. 10, no. 6, e0125777, pp. 1–16, Jun. 2015.

[13] P. Samartsidis, S. R. Seaman, A. M. Presanis, M. Hickman, and D. D. Angelis, "Assessing the causal effect of binary interventions from observational panel data with few treated units," *Statistical Science*, vol. 34, no. 3, pp. 486–503, Aug. 2019.

[14] K. H. Brodersen, F. Gallusser, J. Koehler, N. Remy, and S. L. Scott, "Inferring causal impact using Bayesian structural time-series models," *The Annals of Applied Statistics*, vol. 9, no. 1, pp. 247–274, 2015.

[15] S. Löwe, D. Madras, R. Zemel, and M. Welling, "Amortized causal discovery: Learning to infer causal graphs from time-series data," Feb. 2022, arXiv:2006.10833 [cs.LG].

[16] D. Mønster, R. Fusaroli, K. Tylén, A. Roepstorff, and J. F. Sherson, "Causal inference from noisy time-series data – testing the convergent cross-mapping algorithm in the presence of noise and external influence," *Future Generation Computer Systems*, vol. 73, pp. 52–62, Aug. 2017.

[17] K. W. Soo and B. M. Rottman, "Causal strength induction from time series data," *Journal of Experimental Psychology: General*, vol. 147, no. 4, pp. 485–513, 2018.

[18] B. gong Zhang, W. Li, Y. Shi, X. Liu, and L. Chen, "Detecting causality from short time-series data based on prediction of topologically equivalent attractors," *BMC Systems Biology*, vol. 11, no. 128, pp. 141–150, 2017.

[19] M. Vlachos, M. Hadjieleftheriou, D. Gunopulos, and E. Keogh, "Indexing multi-dimensional time-series with support for multiple distance measures," in *KDD'03*, 2003, pp. 216–225.

[20] C. Cassisi, P. Montalto, M. Aliotta, A. Cannata, and A. Pulvirenti, *Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining*. IntechOpen, London, UK, 2012, pp. 71–96.

[21] G. E. Batista, E. J. Keogh, O. M. Tataw, and V. M. Souza, "CID: an efficient complexity-invariant distance for time series," *Data Mining and Knowledge Discovery*, vol. 28, no. 3, pp. 634–669, May 2014.

[22] A. Zielezinski, S. Vinga, J. Almeida, and W. M. Karlowski, "Alignment-free sequence comparison: benefits, applications, and tools," *Genome Biology*, vol. 18, no. 186, pp. 1–17, 2017.

[23] P. Loskot, K. Atitey, and L. Mihaylova, "Comprehensive review of models and methods for inferences in bio-chemical reaction networks," *Frontiers in Genetics*, vol. 10, no. 549, pp. 1–29, 2019.

[24] V. Wolf, R. Goel, M. Mateescu, and T. A. Henzinger, "Solving research article the chemical master equation using sliding windows," *BMC Systems Biology*, vol. 4, no. 42, pp. 1–19, 2010.

[25] D. T. Gillespie, "Stochastic simulation of chemical kinetics," *Annual Review of Physical Chemistry*, vol. 58, pp. 35–55, 2007.

[26] D. J. Klinke II and S. D. Finley, "Timescale analysis of rule-based biochemical reaction networks," *Biotechnology Prog.*, vol. 28, no. 1, pp. 33–44, 2012.

[27] A. Gupta and P. Mendes, "An overview of network-based and -free approaches for stochastic simulation of biochemical systems," *Computation*, vol. 6, no. 1, pp. 1–20, 2018.

[28] D. Barua, W. S. Hlavacek, and T. Lipniacki, "A computational model for early events in b cell antigen receptor signaling: analysis of the roles of lyn and fyn," *J. Immunology*, vol. 189, no. 2, pp. 646–658, Jul. 2012.

[29] L. A. H. et al., "Bionetgen 2.2: advances in rule-based modeling," *Bioinformatics*, vol. 32, no. 21, pp. 3366–3368, Nov. 2016.

[30] "BioNetGen source code," https://github.com/RuleWorld/bionetgen, accessed: January 2022.

[31] M. W. Sneddon, J. R. Faeder, and T. Emonet, "Efficient modeling, simulation and coarse-graining of biological complexity with NFsim," *Nature Methods*, vol. 8, no. 2, pp. 177–183, Feb. 2011.

[32] "NFsim source code," https://github.com/msneddon/nfsim, accessed: January 2022.

[33] "Stumpy python library," https://stumpy.readthedocs.io/, accessed: January 2022.

[34] Y. P. Park and M. Kellis, "CoCoA-diff: counterfactual inference for single-cell gene expression," *Genome Biology*, vol. 22, no. 228, pp. 1–23, 2021.

 ISBN: 978-1-61208-970-6