Distinguishing Causal and Acausal Temporal Relations

Kamran Karimi and Howard J. Hamilton

Department of Computer Science
University of Regina
Regina, Saskatchewan
Canada S4S 0A2
{karimi,hamilton}@cs.uregina.ca

Abstract. In this paper we propose a solution to the problem of distinguishing between causal and acausal temporal sets of rules. The method, called the Temporal Investigation Method for Enregistered Record Sequences (TIMERS), is explained and introduced formally. The input to TIMERS consists of a sequence of records, where each record is observed at regular intervals. Sets of rules are generated from the input data using different window sizes and directions of time. The set of rules may describe an instantaneous relationship, where the decision attribute depends on condition attributes seen at the same time instant. We investigate the temporal characteristics of the system by changing the direction of time when generating temporal rules to see whether a set of rules is causal or acausal. The results are used to declare a verdict as to the nature of the system: instantaneous, causal, or acausal.

1. Introduction

In this paper we introduce the TIMERS method (Temporal Investigation Method for Enregistered Record Sequences). This method is based on a temporal order among the observed values of the attributes. Suitable input consists of a chronologically ordered set of records, where each record contains the values of the attributes, all observed at the same time. An example of such a record is: <x = 1, y = 2, z = 3>. These records are obtained at regular intervals. Here we are not concerned about the relation among individual attributes, such as "x and y are causes of z." Instead, TIMERS judges a ***set of temporal rules*** that involves the values of x and y to predict the value of z, as being either causal or acausal. An example rule that could belong to this rule set is: **if** {(x = 1) **and** (y = 2)} **then** (z = 3). x and y are considered to be condition attributes, while z is the decision attribute. We cannot tell if this rule, considered by itself, represents a causal relation or not, i.e., do x and y cause z to have a certain value, or do they just happen to be seen together, with all their values caused by some hidden variable(s)? The TimeSleuth software [2] implements the TIMERS method and tries to answer this question. This method is especially appropriate when we have access to many attributes of a system, because the more attributes we have, the better the chance of finding a meaningful relationship among them.

A popular method for assessing the causality of a relationship is to use the concept of conditional independence to determine how two attributes influence each other [8]. In previous work, we looked at other methods of discovering causality, such as TETRAD [10] and CaMML [6]. One important property that differentiates the method we will present in this paper from these is that the presented method deals with data originating from the same source over time, while others deal with data generated from different sources with no special temporal ordering. Our previous work [3,4] was concerned with discovering temporal rules, with no consideration of causality and acausality. Here we discuss an extension to our method to aid a domain expert in making that distinction.

The rest of the paper is organized as follows. Section 2 defines the two directions for time, forward and backward, describes an operation called flattening, and formally defines causality and acausality in the context of the TIMERS method. The distinction among temporal and atemporal rules is also made clear. Section 3 explains how TIMERS determines the nature of a set of rules. Section 4 presents the results of experiments performed with the TimeSleuth software using real and synthetic data sets. Section 5 concludes the paper.

2. Forward and Backward Directions of Time

We consider a set of rules to define a relationship among the condition attributes and the decision attribute. A temporal rule is one that involves variables from times different than the decision attribute's time of observation. An example temporal rule is:

If {(At time T_{-3}: $x = 2$) **and** (At time T_{-1}: $y > 1, x = 2$)} **then** (At time T: $x = 5$). (Rule 1).

This rule indicates that the current value of x (at time T) depends on the value of x, 3 time steps ago, and also on the value of x and y, 1 time step ago. We use a preprocessing technique called flattening [3] to change the input data into a suitable form for extracting temporal rules with tools that are not based on an explicit representation of time. With flattening, data from consecutive time steps are put into the same record, so if in two consecutive time steps we have observed the values of x and y as: Time n: $<x = 1, y = 2>$, Time $n + 1$: $<x = 3, y = 2>$, then we can flatten these two records to obtain <Time $T - 1$: $x_1 = 1$, $y_1 = 2$, Time T: $x_2 = 3$, $y_2 = 2$>. The "Time <number>" keywords are implied, and do not appear in the records. The initial temporal order of the records is lost in the flattened records, and time always starts from ($T - w - 1$) inside each flattened record, and goes on until T. Time T signifies the "current time" which is relative to the start of each record. Such a record can be used to predict the value of either x_2 or y_2 using the other attributes. Since we refrain from using any condition attribute from the current time, we modify the previous record by omitting either x_2 or y_2.

In the previous example we used *forward flattening*, because the data is flattened in the same direction as the forward flow of time. We used the previous observations to predict the value of the decision attribute. The other way to flatten the data is *backward flattening*, which goes against the natural flow of time. Given the two previous example records, the result of a backward flattening would be < Time T: $y_1 = 2$, Time $T + 1$: $x_2 = 3$, $y_2 = 2$>. Inside the record, time starts at T, and ends at ($T + w - 1$). This record could be used to predict the value of y_1 based on the other attributes. x_1 is omitted because it appears at the same time as the decision attribute y_1. In the backward direction, *future* observations are used to predict the value of the decision attribute.

Given a set of N temporally ordered observed records $\boldsymbol{D} = \{\boldsymbol{rec}_1, \ldots, \boldsymbol{rec}_N\}$, the problem is to find a set of rules, as described in more detail below. Each record $\boldsymbol{rec}_t = <c_{t1}, \ldots, c_{tm}>$ gives the values of a set of variables $V = \{v_1, \ldots, v_m\}$ observed at time step t. . The *forward window set* $P_f(w, t) = \{d_t, c_{ki} \mid (w \leq t)$ & $t\text{-}w+1 \leq k < t, 1 \leq i \leq m\}$ represents all observations in the window of size w, between time $(t - w + 1)$ and time t, inclusive, where t is the current time. Time flows forward, in the sense that the decision attribute appears at the end (time t). d_t is the decision attribute at time t. The *backward window set* $P_b(w, t) = \{ d_t, c_{ki} \mid (t \leq |\boldsymbol{D}| - w + 1)$ & $t < k \leq (t + w - 1), 1 \leq i \leq m\}$ represents all observations in the window between time t and time $(t + w - 1)$, inclusive. Time flows backward, in the sense that the decision attribute appears at the beginning (time t). At the time step containing the decision attribute, condition attributes do not appear. In other words, d_t is the only variable at current time t.

Formally, the flattening operator $F(\boldsymbol{w}, \boldsymbol{D}, \textbf{\textit{direction}}, \boldsymbol{d})$ takes as input a window size w, the input records $\boldsymbol{D}$, a time direction ***direction***, and the decision attribute $\boldsymbol{d}$, and outputs flattened records according to the algorithm in Figure 1.

```
for ($t = 1$ to $|\boldsymbol{D}|$)
        if ((direction = forward) and ($t \geq \boldsymbol{w}$))
                output ($\boldsymbol{z} = <z_{ki} \mid d_{pi} \in P_f(\boldsymbol{w}, t)$ & $k = w\text{-}1\text{-}t+p$ & $z_{ki} = d_{pi}>$)
        else if ((direction = backward) and ($|\boldsymbol{D}| - \boldsymbol{w} + 1 \geq t$))
                output ($\boldsymbol{z} = <z_{ki} \mid d_{pi} \in P_b(\boldsymbol{w}, t)$ & $k = p - t$ & $z_{ki} = d_{pi}>$)
```

Figure 1.The flattening operation. The decision attribute d is used by the $P_f()$ and $P_b()$ sets.

The flattened record contains the neighboring w records in the appropriate direction of time. The F_w operator renames the time index values so that in each record, time is measured relative to the start of that record. In each flattened record, the time index ranges from 0 to w-1. The flattened records are thus independent of the time variable t.

Each rule r, generated from these flattened records, is a pair. The first member of a rule is a set of tests. The other member of the rule is the value that is predicted for the decision variable at time 0 or w-1. $r = (Tests_r, d_{val})$, where $Tests_r = \{ Test = (a, x, Cond) \}$, where $a \in V$, x is the time in which the variable a appears, and *Cond* represents the condition under which *Test* succeeds. One example is: $a_x > 5$. d_{val} is the value predicted for d_t (the decision attribute at time t). The CONDITION operator yields the set of variables that appear in the condition side of a rule, i.e., CONDITION(r) = $\{a_x \mid (a, x, Cond) \in Tests_r\}$. Similarly, we define DECISION(r) = $\{d_{\{0,\, w\text{-}1\}}\}$.

There is no consensus on the definitions of terms like causality or acausality. For this reason, we provide our own definitions here. In previous research, we detected sets of temporal rules and assigned the task of whether such a relationship is causal to a domain expert [4]. Here we provide a way to make such distinction. Even though TIMERS provides an algorithmic method for making a decision through a set of metrics, a domain expert must still make the final decision.

2.1 Instantaneous. An *instantaneous* set of rules is one in which the current value of the decision attribute in each rule is determined solely by the current values of the condition attributes in each rule [11]. An instantaneous set of rules is an *atemporal* one. Another name for an instantaneous set of rules is a (atemporal) *co-occurrence*, where the values of the decision attribute are associated with the values of the condition attributes.

Instantaneous definition: For any rule r in rule set R, if the decision attribute d appears at time T, then all condition attributes should also appear at time T, i.e.,
R is instantaneous iff ($\forall\ r \in\ R$, if d_T = DECISION(r), then $\forall\ a_{\mathrm{t}} \in$ CONDITION(r), $t = T$).

2.2 Temporal. A *temporal* set of rules is one that involves attributes from different time steps. A temporal set of rules can be causal or acausal.

Temporal definition: For any rule r in the rule set R, if the decision attribute appears at time T, then all condition attributes should appear at time $t \neq T$, i.e.,
R is temporal iff ($\forall\ r \in\ R$, if d_T = DECISION(r), then $\forall\ a_{\mathrm{t}} \in$ CONDITION(r), $t \neq T$).

We now define the two possible types of a temporal rule:

2.2.1 Causal. In a *causal* set of rules, the current value of the decision attribute relies only on the previous values of the condition attributes in each rule [11].

Causal definition: For any rule r in the rule set R, if the decision attribute d appears at time T, then all condition attributes should appear at time $t < T$, i.e.,
R is causal iff ($\forall\ r \in\ R$, if d_T = DECISION(r), then $\forall\ a_{\mathrm{t}} \in$ CONDITION(r), $t < T$).

2.2.2 Acausal. In an *acausal* set of rules, the current value of the decision attribute relies only on the future values of the condition attributes in each rule [7].

Acausal definition: For any rule r in the rule set R, if the decision attribute d appears at time T, then all condition attributes should appear at time $t > T$. i.e.,
R is acausal iff ($\forall\ r \in\ R$, if d_T = DECISION(r), then $\forall\ a_{\mathrm{t}} \in$ CONDITION(r), $t > T$).

All rules in a causal rule set have the same direction of time, and there are no attributes from the same time as the decision attribute. This property is guaranteed simply by not using condition attributes from the same time step as the decision attribute, and also by sorting the condition attributes in an increasing temporal order, until we get to the decision attribute. The same property holds for acausal rule sets, where time flows backward in all rules till we get to the decision attribute. Complementarily, in an instantaneous rule set, no condition attribute from other times can ever appear. The TIMERS methodology guarantees that all the rules in the rule set inherit the property of the rule set in being causal, acausal, or instantaneous.

3. The TIMERS Method

The TIMERS method is based on finding classification rules to predict the value of a decision attribute using a number of condition attributes that may have been observed at different times. We extract different sets of rules to predict the value of a condition attribute based on different window sizes and different directions for the flow of time. The quality of the set of rules determines whether the window size and time direction are appropriate. We choose either the training accuracy or the predictive accuracy of the set of rules as the metric for the quality of the rules, and the appropriateness of the window size that was used to generate the rules. TIMERS is presented in Figure 2.

We use ε subscripts in the comparison operators to allow TIMERS to ignore small differences. We define $a >_\varepsilon b$ as: $a > b + \varepsilon$, and $a \geq_\varepsilon b$ as $a \geq b + \varepsilon$. The value of ε is determined by the domain expert.

Input: A sequence of temporally ordered data records D, minimum and maximum flattening window sizes α and β, where $\alpha \leq \beta$, a minimum accuracy threshold Ac_{th}, a tolerance value ε, and a decision attribute d. The attribute d can be set to any of the observable attributes in the system, or the algorithm can be tried on all attributes in turn.

Output: A verdict as to whether the system behaves in a instantaneous, causal or acausal manner when predicting the value of a specified decision attribute.

RuleGenerator() is a function that receives input records, generates decision rules, and returns the training or predictive accuracy of the rules.

```
TIMERS($D$, $\alpha$, $\beta$, $Ac_{th}$, $\varepsilon$, d)
   $ac_i$ = RuleGenerator($D$, $d$);  // instantaneous accuracy. window size = 1
   for ($w = \alpha$ to $\beta$)
        $ac_{fw}$ = RuleGenerator($F(w, D, forward, d), d$)   // causality  test
        $ac_{bw}$ = RuleGenerator($F(w, D, backward, d), d$)  // acausality test
    end for

   $ac_f = \max(ac_{f\alpha}, \ldots, ac_{f\beta})$     // choose the best value
   $ac_b = \max(ac_{b\alpha}, \ldots, ac_{b\beta})$

   if $((ac_i < Ac_{th}) \wedge (ac_f < Ac_{th}) \wedge (ac_b < Ac_{th}))$ then discard results and stop. // not enough info

   if $((ac_i >_\varepsilon ac_f) \wedge (ac_i >_\varepsilon ac_b))$ then verdict = "the system is instantaneous"
   else if $(ac_f \geq_\varepsilon ac_b)$ then verdict = "the system is acausal"
   else verdict = "the system is causal"

   if $w_1 = w_2$ and verdict $\neq$ "the system is instantaneous" then
        verdict = verdict + "at window size" + $w_1$

   verdict = "for attribute $d$, " + verdict
   return verdict.
```

Figure 2. The TIMERS algorithm, performed for the decision attribute *d*.

4. Experimental results.

We use three data sets: synthetic artificial life data from a simulated world, a Louisiana weather database, with seven relevant condition attributes, and a Helgoland weather database with three relevant condition attributes.

Series 1: The first series of experiments used data set from an artificial life program called URAL [13]. In a two-dimensional world, a robot moves around randomly, and records its current location plus the action that will take it to the next (*x*, *y*) position. It also records the presence of food at each location. We first set the position along the *x* axis as the decision attribute. In this world, the current *x* value depends on the previous *x* value and the previous movement direction, and is perfectly predictable. The results appear in Table 1(a). The system is not instantaneous, because a window size of 1 (current time) gives relatively poor results. Rather, the system is causal, because the forward test gives the best results. The same conclusions are obtained for the *y* values.

Next we set food as the decision attribute. In this simple world, the presence of food is associated with its position. Because the location of the food does not change, the presence of food can be predicted using only the robot's previous position at a neighboring location and the robot's previous action. As shown in Table 1(b), the system is instantaneous, because a window size of 1 gives relatively good results. We expect about the same results with any other value for the window size.

(a)

Window	Causality	Acausality
1	**46.0%**	
2	**100%**	70.6%
3	100%	71.7%
4	100%	72.8%
5	100%	**74.5%**
6	100%	73.7%
7	100%	73.3%
8	100%	73.6%
9	100%	72.6%
10	100%	73.9%

(b)

Window	Causality	Acausality
1	**99.5%**	
2	99.3%	99.1%
3	99.3%	99.3%
4	99.3%	99.4%
5	99.4%	99.5%
6	99.4%	**99.6%**
7	99.4%	99.5%
8	**99.5%**	99.5%
9	99.3%	99.6%
10	99.3%	99.6%

Table 1.URAL data.(a) Decision attribute is *x*. (b) Decision attribute is presence of food

Series 2: The second experiment was done on a real-world data set, comprising hourly Louisiana weather observations [12]. The observed attributes consist of air temperature, amount of rain, maximum wind speed, average wind speed, wind direction, humidity, solar radiation, and soil temperature. We used 343 consecutive observations to predict the value of the soil temperature attribute. The results are given in Table 2(a). The system is not instantaneous, because a window size of 1 gives poor results. The system is acausal, since the acausality tests gives results as good as or better than those of the causality tests.

Series 3: The next test was done on the Helgoland weather data set [1], which consists of hourly observations of the year, month, day, hour, air pressure, wind direction, and wind speed attributes. Wind speed was selected as the decision attribute. The results from 3000 hours of consecutive observations are given in Table 2(b). With an accuracy threshold of 21% or higher, the data are insufficient to make a judgement. The records are not rich enough to allow TimeSleuth to create rules that can reliably predict the value of the decision attribute. We expect that additional records would not change this.

(a)

Window	Causality	Acausality
1	27.7%	
2	82.7%	75.1%
3	**86.8%**	**87.1%**
4	84.4%	84.7%
5	86.7%	82.9%
6	77.5%	81.4%
7	79.5%	79.8%
8	80.7%	79.8%
9	77.9%	77.3%
10	79.2%	74.0%

(b)

Window	Causality	Acausality
1	**18.9%**	
2	**17.7%**	**20.7%**
3	14.7%	17.2%
4	14.2%	16.9%
5	13.9%	14.5%
6	14.0%	15.2%
7	13.4%	15.0%
8	13.2%	14.9%
9	12.2%	13.9%
10	12.0%	14.7%

Table 2. (a) Louisiana data, *d* = Soil temperature. (b) Helgoland data, *d* = Wind speed.

5. Concluding Remarks

We introduced the TIMERS method for distinguishing causal and acausal sets of temporal rules. We implemented TIMERS in the TimeSleuth software and applied it to three data sets. TimeSleuth correctly categorized two rules sets as causal and another as acausal for an Artificial Life data set, categorized a rule set for the Louisiana weather data set as acausal, and said data were insufficient to make a conclusion about the Helgoland dataset. TimeSleuth is available at http://www.cs.uregina.ca/~karimi/downloads.html.

References

1. Hoeppner, F., Discovery of Temporal Patterns: Learning Rules about the Qualitative Behaviour of Time Series, *Principles of Data Mining and Knowledge Discovery* (*PKDD'2001*), 2001.
2. Karimi, K., and Hamilton, H.J., TimeSleuth: A Tool for Discovering Causal and Temporal Rules, *14th IEEE International Conference On Tools with Artificial Intelligence (ICTAI'2002)*, Washington DC, USA, November 2002, pp. 375-380.
3. Karimi K., and Hamilton H.J., Learning With C4.5 in a Situation Calculus Domain, *The Twentieth SGES International Conference on Knowledge Based Systems and Applied Artificial Intelligence (ES2000)*, Cambridge, UK, December 2000, pp. 73-85.
4. Karimi K., and Hamilton H.J., Discovering Temporal Rules from Temporally Ordered Data, *The Third International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'2002)*, Manchester, UK, August 2002, pp. 334-338.
6. Kennett, R.J., Korb, K.B., and Nicholson, A.E., Seabreeze Prediction Using Bayesian Networks: A Case Study, *Proc. Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'01)*. Hong Kong, April 2001.
7. Krener, A. J. Acausal Realization Theory, Part I; Linear Deterministic Systems. *SIAM Journal on Control and Optimization*. 1987. Vol 25, No 3, pp. 499-525.
8. Pearl, J., *Causality: Models, Reasoning, and Inference*, Cambridge University Press. 2000.
10. Scheines, R., Spirtes, P., Glymour, C. and Meek, C., *Tetrad II: Tools for Causal Modeling*, Lawrence Erlbaum Associates, Hillsdale, NJ, 1994.
11. Schwarz, R. J. and Friedland B., *Linear Systems*. McGraw-Hill, New York. 1965.
12. http://typhoon.bae.lsu.edu/datatabl/current/sugcurrh.html. Content varies.
13. http://www.cs.uregina.ca/~karimi/downloads.html/URAL.java