The AAAI-17 Workshop on
Knowledge-Based Techniques for Problem Solving and Reasoning
WS-17-12

On Inductive Learning of Causal Knowledge for Problem Solving

Seng-Beng Ho **Fiona Liausvia**
Institute of High Performance Computing, Singapore
1 Fusionopolis Way, #16-16 Connexis, Singapore 138632
hosengbeng@gmail.com fiona.liausvia@gmail.com

Abstract

Causal learning is an inductive process and causal knowledge about the world is of paramount importance for intelligent systems, natural or artificial. Given an observation of events happening in the world, how does an intelligent system establish the causalities between them? The issue is further complicated by intervening noisy events. Psychologists have proposed a contingency model of causal induction but it does not incorporate computational means of addressing the issues of intervening noise to recover the causalities between events. In this paper we propose an inductive causal learning method that is able to establish causalities between events in the presence of intervening noisy events, and we apply the method to real-world data to investigate its viability. We demonstrate that the learning method works well in uncovering valid causalities, and relatively non-noisy, opportunistic situations provide the best confirmation of the causalities involved. Causal knowledge is the foundation of problem solving and the ability to learn causal knowledge enables the intelligent system to be maximally adaptive.

Introduction

One of the most important issues for AI specifically and cognitive science in general is the learning of causality. Using the psychologist Patricia Cheng's words, the determination of causality is basically about answering the question: "How does a reasoner know that one thing causes another?" (Cheng 1997). This lies at the foundation of our knowledge building process to gain knowledge about the world and hence the ability to successfully operate in it. However, the issue is not an easy one to tackle as the process is *inductive*. There are techniques that have been developed for *deductive* causal inference, such as the Bayesian causal inference method, which involves inferring the likely causes of certain observed effects by calculating the *a posterior* probabilities involved given the *a priori* probabilities and likelihoods (Pearl 2009). However, the process of deriving the likelihoods requires first having the knowledge of what events (causes) are causally linked to what other events (effects), like answering the Cheng's question above, before the probabilities of these causally linked pairs of events could be calculated to obtain the likelihoods.

In various studies of causal induction, psychologists have shown that often prior knowledge is needed to assist in the process (e.g., Cheng 1997, Griffiths and Tenenbaum 2009, Johnson and Keil 2014). However, humans and animals arrive in this world without any pre-knowledge of anything (though they may have some built-in reflexes of sorts), and must learn a host of causalities for survival. AI systems are also required to have the same ability to learn causalities for maximum adaptability. The process is one of bootstrapping. Firstly, some basic causalities are inductively learned. Then, these form the pre-knowledge to facilitate the further learning of causalities, like in the cases studied by the psychologists. The question addressed in this paper is knowledge-free inductive causal learning. We describe an inductive causal learning mechanism that is able to observe the temporal correlations between events and establish likely causal relationships between them, in the presence of other "noise events." The algorithm is tested on real-world data – specifically the learning of the lightning-thunder causality from videos of real world situations – and the results are presented and discussed.

Note that we are not claiming that temporal correlation is the only way to establish causality, but it can establish some causalities, such as the lightning-thunder causality. These are termed the "ground level" causalities, and other "higher, knowledge level" causalities can be built on them.

Fire and Zhu (2015) had also investigated the inductive recovery of causality from perceptual input. However, their method relies on setting a fixed time window within which to observe putative causal relations. Our method does not impose this constraint.

Basic Considerations

Figure 1 shows a sequence of events, EV1, EV2, … in some temporal order. Suppose an observer-agent with no prior knowledge makes observations on the events as they occur. After observing the second event, EV2, it is reasonable to postulate that perhaps the first event, EV1, is the cause of EV2 based on the temporal connection (assuming that the events are not spaced too far apart temporally – the quantitative consideration of this will be provided later). After the third event, EV3, is observed, there are several possibilities. It could be EV2 that is the cause of EV3: EV2 → EV3; it could be EV1→EV3, with EV2 being "noise"; or it could be that both EV1 and EV2 are required to cause EV3: EV1 & EV2 → EV3. Therefore, at the point after EV3, we could postulate all 4 possibilities, including the earlier EV1→EV2, and these can be placed on a "tentative list of possible causalities" (TLPC). Further observation is needed to discern which are "true" causalities and which are "noise." (On the issue of whether "correlation implies causality," Ho (2014, 2016) provides an extensive discussion on this issue and proposes the idea of "effective causality." Hence, here we take any good, consistent temporal correlation to imply there is "causality" between the event in terms of *effectively* using the information for problem solving and other inference processes.)

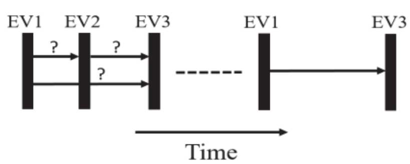

Figure 1. A sequence of events and possible causalities.

One of two situations could obtain after this point in time. There could be more and more different events that happen and this would increase combinatorially the number of possible cause-effects relations to be kept on the TLPC. Another situation is shown in Figure. 1 in which some of the earlier events happen again that provide information on what the likelier candidates are for causality on the TLPC. For example, if EV1 and EV3 happen again without an intervening EV2, it is likely then that 3 out of 4 of the cause-effect relations on the current TLPC, i.e., those that involve EV2, could be discounted. Of course, due to noise (e.g., the unreliability in observation, or that EV2 is just probabilistically associated with the other events), EV2 may just be fortuitously missing. Therefore, the cause-effect relations that involve EV2 should not be totally removed from the TLPC list at this point, instead, some measure of confidence associated with them could be reduced, and further observation would either increase or decrease these confidence measures.

To reduce the possibility of combinatorial explosion of TLPC, we propose instead to just record the adjacent event pairs on the TLPC. I.e., only EV1→EV2 and EV2→EV3 are recorded after the observation of EV3. The TLPC would then grow linearly as more events are observed. One may then ask, what happens to the possible relations EV1→EV3 and EV1&EV2 → EV3? For EV1→EV3, if indeed it is the case that EV2 is noise, more EV3 following EV1 without the intervening EV2 would be observed, and that would establish EV1→EV3 as a strong causal relation later. As for EV1 & EV2→EV3, if indeed EV1 and EV2 are conjunctively needed for the occurrence of EV3, then EV1→EV2 and EV2→EV3 would both emerge as strongly temporally correlated, and this would be equivalent to EV1 & EV2 → EV3.

The Inductive Causal Learning Algorithm

In Figure 2 we use some well-known day-to-day events to demonstrate the proposed inductive causal learning algorithm. The algorithm is an "online" algorithm as it establishes causalities as observation reveals events across time. In the world we know, lightning is the cause of thunder. We know this perhaps first from sensory information, that lightning seems always to precede thunder. After humanity has gained a deeper understanding of physical processes, we know that thunder is caused by the disturbance to the air as a result of electricity traveling through it, which is lightning. But presumably way before humanity has understood these physical processes, thunder has already always been known to be the effect of lightning, and not the other way around, and also not as an effect of other events (such as someone sneezing just prior to the thunder). That knowledge can only be obtained from observing the temporal correlations from direct sensory observation, and also in the presence of noise such as other preceding, succeeding, or intervening events between lightning and thunder. Therefore, in Figure 2 we include other events such as wind (W), headlight of cars (H), and sound of cars or other objects (S).

Firstly, we consider the event pairs of interest with no intervening noise. In Figure 2 we show that sometimes lightning is followed immediately by thunder but sometimes thunder is missing. In the case when thunder is missing, it could be because there is intervening noise and it may appear later, or it could be because it is "really" missing in the sense that even after observing some other intervening events, it is still not observed and instead lightning is observed – this could be due to a "weak" lightning that somehow does not generate thunder or that the observation of thunder fails because the thunder is too soft or being obscured by other sounds.

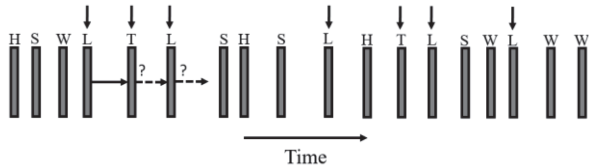

Figure 2: Lightning (L) and thunder (T) events, along with events that intervene between them – wind (W), car headlight (H), car or other sound (S). L and T events are highlighted by vertical arrows.

We define Forward Immediate Probability, *FIP*, as the probability of observing an effect given a certain putative cause. Therefore, after the second lightning event in Figure 2, the *FIP* for the *Lightning→Thunder* relation is 0.5 as thunder fails to appear immediately after the second lightning. However, as we mentioned above, thunder may appear after "skipping over" some other intervening events. This is *FSP* – forward skipped over probability. Therefore, after the last lightning event in Figure 2, the *FSP* for the *Lightning→Thunder* relation is 0.33. This is because out of the 3 event sequences, L—S—H—S—L, L—H—T, and L—S—W—L, in which lightning happens first followed by some other intervening events, only one, the L—H—T sequence, ends with a thunder.

In a similar manner, we define *BIP*, backward immediate probability and *BSP*, backward skipped over probability. *BIP* is the probability of observing the immediately preceding event – i.e., for thunder, how often lightning preceeds it. *BSP* is the backward probability considering skipping over other events.

We define the probability of EV1 causing EV2 as follows:

$$Prob\ (Cse(EV1, EV2)) = FIP+FSP-(1-(BIP+BSP)) \quad (1)$$

That is, the total of the "forward" probabilities, *FIP+FSP*, increases the likelihood that there is a causal relation between EV1 and EV2 and the total 'backward" probabilities provides a "check" on the forward probabilities – i.e., if the total backward probabilities is 1, then it means the cause consistently precedes the effect, and that means the total of the forward probabilities, *FIP+FSP*, is not being reduced at all by the remaining terms. Otherwise, the smaller the *BIP+BSP*, the less likely EV1 and EV2 are causally linked.

At this point, we would like to pause and compare the above equation to that proposed in a so-called contingency model of causal induction used in psychology (Rescorla 1968, Jenkins and Ward 1965). The formula is stated as follow:

$$\Delta P_i = P(e \mid i) - P(e \mid \neg\, i) \quad (2)$$

ΔP_i is the "contingency" between a candidate cause (event) *i* and an effect (event) *e*. $P(e \mid i)$ is the probability that effect *e* happens given that the cause *i* has happened, and this probability is reduced by the probability that e happens without *i* having happened, $P(e \mid \neg\, i)$, to derive the contingency ΔP_i. Eqn. 1 parallels exactly Eqn. 2 except that we separate the corresponding probabilities into "immediate" and "skipped-over" probabilities, thereby incorporating the consideration of "noisy" intervening events. Our method is therefore a practical implementation of Eqn. (2) with consideration of "noisy" intervening events.

The consideration of causally linked events in the presence of skipped-over events allows noise to be bypassed, but this brings back the issue of combinatorial explosion discussed in connection with Figure 1. We therefore define a threshold called Skipped-Over Consideration Threshold (SOCT) that is used to activate skipped-over considerations only when the putative causality has acquired some degree of likelihood. We use a certain percentage (e.g., 30%) of the highest value of *FIP* (corresponding to a certain cause-effect pair) as the SOCT and include all the putative cause-effect pairs that have higher *FIPs* than the SOCT for skipped-over consideration (i.e., their *FSP* and *BSP* would be computed). In a situation in which there are not many different kinds of events, which means there will not likely be combinatorial explosion of the TLPC list, it is alright to set SOCT to 0 so that good cause-effect relations will not be missed.

Building on the Basic Causal Formula – the Issue of Asymmetry

Having established a basic formula describing the putative causal relations between two events, there are two other issues that need to be considered that has bearing on the "strength" of the causal relation. One is the issue of temporal symmetry between a pair of events and the other is the amount of noise that intervenes between them.

Figure 3 shows a sequence of an event EV1 alternating with another event EV2. Considering a real-world example, EV1 could be lightning and EV2 could be thunder. If one observes a sequence such as this, how could one distinguish whether it is EV1 that causes EV2 or the other way around? Eqn. 1 or 2 does not provide any means to exploit the asymmetry between EV1 and EV2 to provide a means for the discernment of the causal direction.

EV1—EV2—EV1—EV2—EV1—EV2…

Figure 3: Two alternating events, EV1 and EV2.

In some kinds of event pairs such as lightning (EV1) and thunder (EV2), there is an asymmetry between the intervals EV1—EV2 and EV2—EV1. Typically, the last thunder may have ended many days, weeks, or months ago and in a lightning-thunder season, one observes lightning happens "first' after a long hiatus. The causal direction is then more likely to be EV1→EV2 rather than EV2→EV1. In some other kinds of event pairs, such as two alternatively blinking lights with exactly identical EV1—EV2 and EV2—EV1 intervals (such as that might be obtained from an electronic circuit controlling the lights), there is no asymmetry and the causalities in both directions are equally likely.

We define the uncertainty associated with interval length, *UIL*, as follows:

$$UIL(EV1, EV2)= F_1(Av(IL(EV1, EV2))/Av(IL(\forall\ EVP))) \quad (3)$$

where *Av(IL(EV1, EV2))* is the average interval length of the event pairs EV1—EV2, and *Av(IL(∀ EVP))* is the average interval length of all event pairs, including EV1—EV2. The purpose of using the quotient between these two quantities to create a relative measure of interval length is that whether an interval is considered large or small is relative to what other intervals are like in absolute terms. F_1 is a function that bounds *UIL* between 0 and 1 so that it has a comparable magnitude to the probability measure of Eqn. 1 in order that they can be combined in some manner in a formula to be described below. F_1 is defined as:

$$F_1(x) = 0 \text{ for } x \leq 1$$

$$F_1(x) = 2/(1+EXP(-k(x-1))) - 1 \text{ for } x > 1 \quad (4)$$

where k is a quantity that determines the rate of change of the exponent. F_1 is derived from the logistic function. For $k = 0.2$, when x is 10, which means when the interval length involved (for an event pair) is 10 times that of the average interval length of all event pairs, the value of F_1 is 0.72, which is quite close to the "worst" value of 1. When the value of x is 1, F_1 is 0, which means that if the average value of the interval involved is the same as that of the average of all the interval lengths associated with all event pairs, the uncertainty in the interval length associated with the event pair involved is 0.

Other than the average value of the interval length, the variability of the interval length is also an indicator of the strength of the causal relationship. If the variability is low, the causality between the events involved is more certain. We define the uncertainty due to interval variability, *UIV*, as follows:

$$UIV(EV1, EV2) = F_2(Dv(IL(EV1, EV2)) / Av(IL(EV1, EV2))) \quad (5)$$

where *Dv(IL(EV1, EV2))* is the deviation of the values of the interval lengths between event pairs EV1—EV2. The deviation can be computed by any means. A simple calculation for *Dv* takes the difference between the largest and smallest interval values. Standard deviation can also be used. F_2 serves the same function as F_1 in Eqn. 4 to bound the value of *UIV* between 0 and 1 except that we desire F_2 to be 0 when x is 0, instead of when x = 1 like in the case of F_1 above, because we want the uncertainty involved to be 0 when the deviation is 0. F_2 is defines as:

$$F_2(x) = 0 \text{ for } x \leq 0$$

$$F_2(x) = 2/(1+EXP(-k(x))) - 1 \text{ for } x > 0 \quad (6)$$

Intervening Noise

Other than interval length, the number of other events that intervene between a pair of events under consideration is also a reflection of the strength of the causal relation between them. So, if lightning is followed by thunder without other intervening events, the *lightning→thunder* causality is "strong." If these "noisy events" are randomly distributed, then the interval length measure described in the previous section is proportional to the measure of intervening noise. However, there can be situations in which they are independent. For example, during a storm, not only there are a lot of lightning—thunder events, there are usually also a lot of wind, rain, fewer or more vehicles plying the roads, etc. We therefore measure the uncertainty in the causal relation arising from intervening noise, *UIN(EV1, EV2)*, as follows:

$$UIN(EV1, EV2)= F_1(Av(NOE(EV1, EV2)) / Av(NOE(\forall\ EVP))) \quad (7)$$

where *NOE(EV1, EV2)* is the number of other intervening events between the event pair EV1—EV2. *Av* and ∀ *EVP* have the same meaning as in Eqn. 3. F_1 is as defined in Eqn. 4. In a similar vein as Eqn. 5, we define the uncertainty due to the variability in the number of intervening other events, *UINV(EV1, EV2)*, as follows:

$$UINV(EV1, EV2) = F_2(Dv(NOE(EV1, EV2)) / Av(NOE(EV1, EV2))) \quad (8)$$

where *Dv* and F_2 are as defined in Eqns. 5 and 6 respectively.

Strength of Causal Relation

With the various uncertainty measures, we can calculate a total uncertainty, *TU*, for the putative causal relation of the event pair EV1—EV2 as follows:

$$TU(Cause(EV1, EV2)) = UIL+UIV+UIN+UINV \quad (9)$$

where *UIL, UIV, UIN, UINV* are as defined in Eqns. 3, 5, 7, and 8 respectively. With this, we define the *Strength* of causality of the event pair EV1—EV2 as:

$$Strength(Cse(EV1, EV2)) = Prob(Cse(EV1, EV2)) - w*TU(Cse(EV1, EV2)) \quad (10)$$

where *w* is use to scale the contribution of *TU* to the *Strength* value.

Testing with Real-World Data

For testing with real-world data, we selected some YouTube videos of lightning and thunder events, which are typically also accompanied by other intervening events. Figure 4 shows a screenshot of one of the videos we selected named "Nonstop Thunder and Lightning!": https://www.youtube.com/watch?v=csODdOOxEpk.

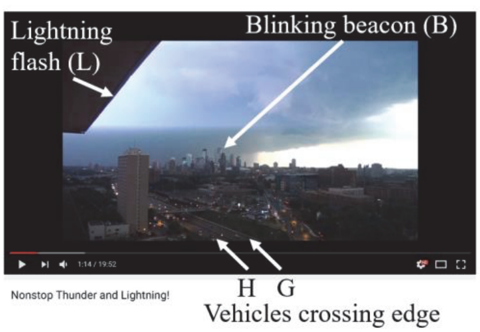

Figure 4. Screenshot of the Nonstop Thunder and Lightning video: https://www.youtube.com/watch?v=csODdOOxEpk.

This video consists of about 20 minutes of recording of a number of lightning and thunder events, together with cars moving in the foreground and a blinking beacon on a building in the distance. The lightning events come mostly in the form of "flashes in the sky" rather than the typical distinct electrical discharges characteristic of lightning. The thunder events can be heard as "rumbles from the sky." In the following we describe the experiments performed on 3 separate videos of lightning and thunder. In all experiments, the SOCT (discussed in connection with Eqn. 1) is set to 0, the deviation, *Dv* (Eqns. 5 and 8), is the average of the sum of the absolute values of the differences between the data values and their average value. K (Eqns. 4 and 6) was set to 0.2 and w (Eqn. 10) was set to 0.5.

The Nonstop Thunder and Lightning Video

The first experiment was performed on the video whose screenshot is shown in Figure 4. Visually, the event pattern stays more or less the same throughout the entire 20-minute length of the video, so we selected at random three 1-minute segments to collect data from to perform 3 tests: Test 1: 3:01 to 4:00 minute, Test 2: 5:01 – 6:00 minute, Test 3: 8:01 – 9:00 minute. The main events that are observable are lightning (L), thunder (T), beacon (B), and moving cars. These events were manually identified and the times at which they happen were manually noted. Because thunder tends to appear as a long-drawn rumble, we take the time of the start of the rumble to represent the time of the thunder event. We use the moving cars to create "headlight observation" events. We assume an observer is located at a specific place to observe the various events, including "flashes of car headlight" shining at her. To simulate this effect, we selected the bottom edge of the video frame at which the cars move out of or into the video frame, indicated in Figure 4 as "vehicles crossing edge," and recorded the moments of crossing to be the moments of "headlight flashes events." There are two streams of traffic, one coming toward the edge and the other going away from it. We label the points of crossing as H and G respectively and these are the points at which the observer observes the "flashes of headlight" events, Hs and Gs.

Figure 5(a) shows the collected data of the five events: T, L, B, H, and G in the Test 1 segment. Each time step in Figure 5(a) represents 200 ms (milliseconds) of time. We work with a temporal resolution of less than 1 second because often more than 1 event happens within 1 second. However, it is difficult for humans to discern the order of events taking place within a 1-second interval. We therefore randomly assign the order of events observed within each 1-second interval.

Figure 5(b) shows the cause-effect *Strengths* of the various event pairs in Figure 5(a) based on Eqn. 10. There are two kinds of *Strength* values computed and displayed in Figure 5(b). They are for the same-event-pairs and different-events-pairs. The same-event-pairs are made up of causes and effects that are both the same kind of events, such as L and L. As mentioned earlier, there could be repeated events, such as blinking lights, that have a regularity and computing their cause-effect *Strengths* would be useful as they would allow us to predict/anticipate the next event (and as mentioned, whether this is "truly causal" does not matter as what we are after is *effective* causality (Ho 2014)). Now, due to the way we compute the *Strength* values, there is a positive "bias" toward giving same-event-pairs much higher values because unlike different-events-pairs, in which sometimes the "cause" event is not followed by the "effect" event and vice versa (with "immediately following/preceding" or skipped-over considerations), thus lowering the *Strength* values involved, there are always an instance of an event following or preceding another instance of that same kind of event in a long sequence of that event, except at the beginning or the end of the sequence. Therefore, the *Prob(Cse(EV1, EV2))* part (Eqn. (1)) of the *Strength* calculation is always high and only the value of the *TU* portion (interval length and variability, etc.) would determine the overall *Strength* value. Hence, the data for same-event-pairs and different-event-pairs are separately considered, and the same-event-pairs are shaded in gray.

Since our focus is on how well can the method uncovers the L→T relation, we evaluate the results based on 4 criteria: (1) How strong is the L→T *Strength* value; (2) By how much do other event pairs' *Strength* values exceed the L→T *Strength* value (excluding same-event-pairs); (3) How different is the L→T Strength value from that of the next strongest one (excluding same-event-pairs); (4) How different is the L→T *Strength* value from that of the T→L value. Following this evaluation, we will discuss, in the Discussion and Conclusion section, how a system can extract *any* meaningful causalities using the current method, in the absence of any pre-knowledge of what causality to look out for.

In Figure 5(b) it can be seen that though the L→T *Strength* is at the top of the list, at a value of 0.357, the value is not high. It is not that well separated from the next higher event-pair H→L even though it is very well separately from T→L in terms of the *Strength* values. The reason why this value is low is that L is not always followed by T and T is not always preceded by L, even including skipped-over considerations, thus lowering the *Prob(Cse(EV1, EV2))* value of Eqn. (1). The *TU* portion of the equation for *Strength* (Eqn. (10)) also contributed to the low overall value because of the values of *UIL*, *UIV*, *UIN*, and *UINV*.

The reason why the Strength value of L→T is not well separated from the H→L value is that in the data stream that we collected, all the events continue to occur with similar frequencies. In principle, cars could have stopped moving, and lightning and thunder could go on, if a longer data stream is collected. That would give rise to a much lower H→L value and L→T would be better separated from it.

As for the high value of the same-event-pair, H→H, it is due to the relative regularity of the H—H events. Therefore, we can say that an H event predicts another H event quite well. For this H—H event pair, we know from background knowledge that it is not "causality" per se, just a stream of headlights from moving cars, but the strong correlation is a measure of how well the observation of one H events allows the observer to predict a subsequent H event.

```
__G_H__B____H__H_B_LT__B______H_B__LTBGH
_H___BH_LG_BH_G_L__G_H_BL___B_B____G___L
T_B__L_G_H_BH_G_HG_B_____LT_GHLGT_H_H__
_LTH_G___H______L_H__L_H_G_G_L_TH_G___H_
_T__H_H___L_H_GL_H________H___GL_H_H_G__
_LT_G_H_G__H___H___G_____GL__B__H___B___
L_TBHTL__GT_B_HLBT_G_H____BH_G_B___H_LT
```

(a)

Cause→ Effect	Strength	Cause→ Effect	Strength
H→H	0.847	B→G	-0.029
L→T	**0.357**	L→H	-0.152
H→L	0.273	T→B	-0.167
L→G	0.245	B→B	-0.169
G→H	0.239	T→H	-0.177
G→L	0.212	T→G	-0.197
B→L	0.137	G→B	-0.201
H→G	0.082	L→B	-0.304
H→B	0.052	G→T	-0.324
B→H	0.040	H→T	-0.458
G→G	-0.020	**T→L**	**-0.682**
		B→T	**-0.799**

(b)

Figure 5. (a) Event data collected from the Test 1 segment. Events begin from top left corner. L=lightning, T=thunder, B=beacon, H,G=headlight flashes as described in text. (b) The result for the Test 1 segment. Same-event-pairs are shaded in gray.

Figure 6 shows the event data and results for the Test 2 segment of the Non Stop Thunder and Lightning video of Figure 4. The results are similar to that of Figure 5 – the *Strength* value of L→T is not high (though higher than before at 0.461), but is the highest (excluding the same-event-pair, G—G) and well separated from that of the T→L pair, though not well separated from the next highest value pair for the same reason as that explained in connection with Figure 5(b).

The event data collected from the Test 3 segment of the Non Stop Thunder and Lightning video shows a similar pattern as that of the Test 1 and 2 segments. This shows that lightning, thunder and other events continue to occur with a similar pattern across time in this video.

In the two sets of results shown in Figures 5(b) and 6(b), L→T does not score high for criterion (1) but scores well for criteria (2) and (4). It does not score well for criterion

(3) for the above reason explained in connection with Figure 5(b) (i.e., comparing H→L and L→T). We conclude that L→ T is moderately successfully recovered from the observation in this set of test data.

```
__G__B_H__GHLT_H_B__T_GBH_____LGH_B_B_G_

_T___BH__G__T_BL_T_G_G__B_____LHTBG_B___L

_H_G__G___H_T__G____L__T___G__G___T___H_H

G_____G_L_T_G__H_G__H___G________LT__H__G

_LG_T__T__H__G_L_GH___T____HLG______G____

L____H_L_GH____G______G___H________H__TGL

_____L___G_________HT_G___L__T_____G_L_T_G

_H_L_H_L____
```

(a)

Cause→Effect	Strength
G→G	0.540
L→T	**0.461**
H→L	0.371
H→G	0.278
T→G	0.259
G→H	**0.248**
L→G	0.172
G→T	0.157
B→B	0.148
L→H	0.112
T→H	0.099
H→H	0.043

(b)

Figure 6. (a) Event data collected from the Test 2 segment. (b) The results of the Test 2 segment. Only causal pairs with positive Strength values are shown.

The Severe Thunderstorm Derecho Video

In the next experiment, we used a "Severe Thunderstorm Derecho" video from YouTube: https://www.youtube.com/watch?v=QPSdgaY1cyU. Visually there does not appear to be a strong correlation between lightning and thunder and the results also scored badly in terms of the above 4 criteria and we conclude that the L→T relation is not successfully recovered in this set of test data.

The Best Lightning Strike Compilation #6 Video

In the last experiment, we used the "Best Lightning Strike Compilation #6" video from YouTube: https://www.youtube.com/watch?v=P7K3m2zHEhs. We collected data in the time period of 1:00-4:30 minutes. The video consists of a compilation of a number of sequences of lightning and thunder events, but the commonality between them is they are all very distinct lightning and thunder events, with distinct electrical discharges. A screenshot of the lightning is shown in Figure 7.

Figure 7. Lightning and thunder from YouTube video: https://www.youtube.com/watch?v=P7K3m2zHEhs. The lightning and thunder events are very distinct.

Even though this video consists of a number of lightning-thunder sequences strung together, we treated them as though they were all from the same sequence. Only lightning and thunder events were collected as there are no significant other kinds of events. The events were collected at 1 second intervals, unlike the 200 ms intervals in previous experiments. The event data collected and results are shown in Figure 8.

```
_______L___L___LT__________L___T________

_______________L__T__________________L_T

___________________________L_T________LT

___________L__T___________LT____________

__L_T____________________T________T______

__________
```

(a)

Cause→Effect	Strength
L→L	0.791
L→T	**0.720**
T→L	**0.610**
L,T→L	0.083
T→T	-0.809

(b)

Figure 8. (a) Event data collected from the Best Lightning Strike Compilation #6 video on YouTube. (b) The results of the event data of (a).

It can be seen in Figure 8(b) that the L→T causality has a very high *Strength* value. There are no other different-events-pairs that have higher *Strength* values than it. However, even though the T→L causality value is lower than it, they are not that well separated. This is due to the fact that in the data stream that we collected the events, lightning and thunder did not stop. If the situation is like that of the real world, in which typically after a last thunder event ended, it may be days or weeks before the first lightning event would occur, then the interval variability of T→L would be very high and the T→L causality would have a lower *Strength* value. Therefore, we consider the L→T causality is well recovered here.

Discussion and Conclusion

From the above experiments we can see that it is indeed possible for the inductive causal learning process described to be applied to real-world situations to recover the lightning—thunder causality, and that the best situation to recover the causality is the situation of Figure 7, which consists of very distinct lighting and thunder causalities with practically no other noisy events. However, this is from an evaluation point of view in which we are looking for a known L→T causality. If we do not have the pre-knowledge of or a way to determine what a good situation is for causal recovery, how do we discover the lightning—thunder causality?

In the real world, all the various situations in the various experiments, from the favorable situation of Figure 7 to the moderately favorable situation of Figure 4 and the very unfavorable situation of the Severe Thunderstorm Derecho Video could all take place one after another, either immediately following one another or after some time gaps. If the algorithm simply collects all lightning and thunder events from the first moment of observation, with or without noisy interventions and with distinct or indistinct lightning—thunder causalities, the low *Strength* L→T episodes may drown out the high *Strength* ones and at the end no distinct L→T causality such as that from the situation of Figure 8 could be established.

An approach for the recovery/establishment of causality could rely on a measure of "favorable" or "opportunistic" situation. An opportunistic situation for a putative pair of causally linked events can be defined as one that produces a high *Strength* value for that pair of events. Any good causalities detected in such a situation can be "locked-in" in that we do not allow any further noisy situations to reduce the *Strength* already established. In fact, we may use the established high *Strength* value, such as that for the L→T causality that is recovered, say from a situation like that in Figure 8, to interpret the data in noisier situations such as that of Figure 4 – e.g., if L is connected with T, perhaps it is less likely that they participate in other causal relations.

Therefore, the inductive causal learning algorithm described in this paper can be used to learn causalities quickly in opportunistic situations, and the earlier learned causalities can assist in further establishment of other causalities in a bootstrapping process.

Knowledge of causality is the foundation of problem solving. For example, if an intelligent system learns that pushing an object causes it to move in a particular manner, it can later use that knowledge for problem solving: If a certain movement is desired for an object, the corresponding action can be effected on it. This way, the knowledge for problem solving is learned from the environment and not built-in and the system involved would be highly adaptive: New causal rules/knowledge can be learned when the environment changes.

Acknowledgement

This research is supported by the Social Technologies+ Programme funded by A*STAR Joint Council Office.

References

Cheng, P. W. 1997. From Covariation to Causation: A Causal Power Theory. *Psychological Review* 104(2): 367-405.

Fire, A. and Zhu, S.-C. 2015. Learning Perceptual Causality from Video. *ACM Transactions on Intelligent Systems and Technology*, 7(2), 23. doi:10.1145/2809782.

Griffiths, T. L. and Tenenbaum, J. B. 2009. Theory-Based Causal Induction. *Psychological Review* 116(4): 661-716.

Ho, S.-B. 2014. On Effective Causal Learning. In *Proceedings of the 7th International Conference on Artificial General Intelligence*, 43-52. Berlin: Springer-Verlag.

Ho, S.-B. 2016. *Principles of Noology: Toward a Theory and Science of Intelligence*. Switzerland: Springer International.

Johnson, S. G. B. and Keil, F. C. 2014. Causal Inference and the Hierarchical Structure of Experience. *Journal of Experimental Psychology: General* 143(6): 2223-2241.

Pearl, J. 2009. *Causality*. Cambridge, UK: Cambridge University Press.