A Ground Level Causal Learning Algorithm

Seng-Beng Ho and Fiona Liausvia
Department of Social and Cognitive Computing
Institute of High Performance Computing, A*STAR
Singapore, Republic of Singapore
hosengbeng@gmail.com, fiona.liausvia@gmail.com

***Abstract*—Open domain causal learning involves learning and establishing causal connections between events directly from sensory experiences. It has been established in psychology that this often requires background knowledge. However, background knowledge has to be built from first experiences, which we term ground level causal learning, which basically involves observing temporal correlations. Subsequent knowledge level causal learning can then be based on this ground level causal knowledge. The causal connections between events, such as between lightning and thunder, are often hard to discern based on simple temporal correlations because there might be noise – e.g., wind, headlights, sounds of vehicles, etc. - that intervene between lightning and thunder. In this paper, we adopt the position that causal learning is inductive and pragmatic, and causal connections exist on a scale of graded strength. We describe a method that is able to filter away noise in the environment to obtain likely causal connections between events.**

Keywords—causal learning; effective causal learning; open domain causal learning; ground level causal learning; knowledge level causal learning

I. Introduction

Being able to establish causality between events is paramount for the adaptability and survival of any intelligent system, natural or artificial. Yet, despite its seeming simplicity, the conditions under which an event is established to be a consistent cause or effect of another event have baffled psychologists and artificial intelligence (AI) practitioners alike. Even though the terms "causality" and "causal inference" have been in widespread use in the literature, in Ho [1] we distinguish closed vs open domain causality. In Bayesian causal inference (e.g., [2]), the conditional probabilities of some events causing some other events are known, and the inference involves determining what the most likely cause is, given the observed effect. We term this closed domain causal inference. Open domain causal inference, on the other hand, involves determining which events are in a causal relationship, before even these probabilities can be computed. In the very beginning of a well-known psychological paper on causality [3], Cheng asks the question "How does a reasoner know that one thing causes another?" This is the issue of open domain causality that the current paper is concerned with. It has been argued that in general, the determination of causality requires background knowledge, and the observation of a simple temporal correlation between two events is not sufficient [3, 4]. However, this begs the question as to how background knowledge is learned in the first place. In this paper we demonstrate that it is possible to carry out what we term "ground level causal learning," which makes use of temporal correlation, in the presence of noise. Higher level causalities can then be built upon the basic causal knowledge learned in this manner.

II. Basic Considerations

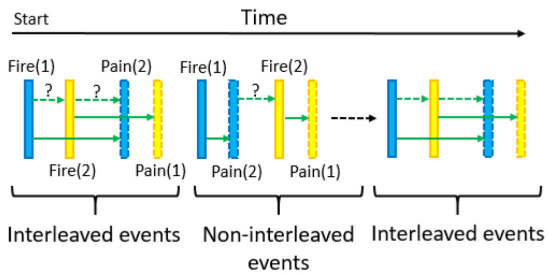
A. Basic Framework

In Ho [1, 5] we discussed the typical problem associated with determining causality from correlation, which is that "correlation does not imply causality." [6, 7] We established that temporal correlation could be "effectively causal" in the sense that if we observe directly and proximally two events in a certain temporal relationship (i.e., one happens after another), we can make use of this correlation to predict what would happen next given the first event, notwithstanding whether there is a third "underlying" event that causes them both. This knowledge is useful for problem solving, even if there exists an underlying third event that is the "true cause." Our approach to causal learning is hence an inductive and pragmatic one. This is the basic framework within which we believe basic causalities about the environment is learned.

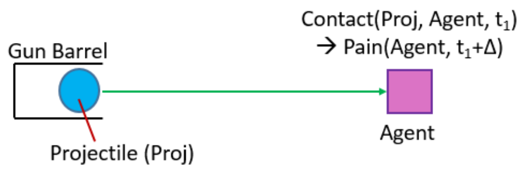
B. Ground Level vs Knowledge Level Causal Learning

In Section I we mentioned that at some level of causal learning, background knowledge is required. However, it was also mentioned that there exists a ground level causal learning on which the "higher" level causal learning is based. Fig. 1 uses a simple example to illustrate this idea.

In Fig. 1(a) it is shown that an observer-agent observes a sequence of firing and pain events that arise due to the firing of a gun, say, and the bullet hurting someone respectively. There are two agents shooting at each other and Fire(1) refers to the firing of the gun by the first agent and Pain(2) refers to the pain experienced by the second agent, and vice versa. The firing event is usually detected as a result of a "bang" sound coming from the gun (or some flashes of light from the barrel if the observer-agent is looking at the gun) and the pain event is observed from the wincing expression of the agents involved (or the observer-agent could be one of the participants of the shooting and she experiences the pain directly).



(a)



(b)

Fig. 1. (a) Firing and Pain events in a shooting scenario. (b) Background knowledge of the working of guns which consists of a projectile (bullet) hurting an agent.

Let us consider first the first set of events in Fig. 1(a) which we label "interleaved events" consisting of two firing events and two pain events. In the interleaved events situation, the cause-effect sequences are not clearly separated as shown. If the observer-agent does not have any background knowledge concerning the causal connection between shooting and pain, it will be difficult for her to discern the "correct" causal relationships – i.e., Fire(1) → Pain(2) and Fire(2) → Pain(1). This is especially true because typically the speed of the bullet is extremely fast and the observer-agent does not observe the intervening physical event between firing and pain. Therefore, Fire(2) may be construed as the effect of Fire(1) and Pain(2) the effect of Fire(2) because they follow the "cause" immediately.

However, suppose the observer has the basic knowledge of a gun as depicted in Fig. 1(b), which depicts a projectile being propelled from a barrel along a trajectory until it hits a recipient agent, she would then be able to understand that the barrel of the gun typically has to be pointing at the recipient agent before the gun can hurt her (there is of course also the possibility of richochet, which obeys certain laws of reflection, which can also be relied on to infer whether a bullet emerging from a certain gun could potentially hurt someone). The basic understanding here involves understanding that the projectile has to contact the recipient agent, and the contact is the *cause* of the hurt (shown in a causal rule Contact(Proj, Agent, t_1) → Pain(Agent, $t_1+\Delta$) in Fig. 1(b)). This understanding could have come from observing a slower, observable projectile movement. This process has been discussed in Ho [1, 8].

Now, interestingly, the learning of this Contact → Pain causal rule can be achieved through a ground level causal learning process much like that depicted in Fig. 1(a), through temporal correlation [1] – at t_1 there is a contact event and then at $t_1+\Delta$ there is a pain event. This could then be packaged in some knowledge structure that encodes the idea of how a gun could cause hurt through a bullet, which in turn then provides an observing-agent with the ability to discern the causality correctly from the observation of the first set of interleaved events. In so doing, the observing-agent would need additional observational information such as the gun barrel's direction to establish the correct pairs of causes and effects.

This is the process through which all causal knowledge is learned – some basic causal knowledge is first learned through the ground level process of temporal correlation, and then as more knowledge is acquired, they could participate in subsequent knowledge level learning process.

Returning to Fig. 1(a), we now consider the second set of events which we label "non-interleaved" events. In this case, the correct cause-effect pairs could be easily discerned, with the effects following the causes immediately.

Now, assuming that a number of non-interleaved events has been observed and the observing-agent has gained some confidence in the right pairings of causes and effects, if she then later encounters some interleaved events such as the second set of interleaved events shown in Fig. 1(a), she can then use this earlier-learned causal knowledge to correctly pair up the causes and effects.

We think of the set of non-interleaved events in Fig. 1(a) as a kind of opportunistic situation that allows the correct cause-effect pairs to be learned, and the correct pairings of the earlier encountered interleaved events can be retroactively established using this knowledge. Opportunistic situations can allow the correct cause-effect pairs to be learned in the absence of background knowledge. In the following, we will base a ground level causal learning process on this idea with consideration of noise – for a given cause-effect pair, the cause or effect may not always take place or be always observable, and there could also be other "noisy" events intervening the cause and effect.

In the set of non-interleaved events in Fig. 1(a), we indicate a possibly spurious cause-effect pair with a "?". In general, someone feeling pain from a bullet from someone else's gun (say, Pain(2)) and her own firing of a gun (say, Fire(2)) are not correlated – if two agents are involved in a gun fight, they would want to keep firing at the others anyway, not as a result of receiving some pain, unless they are observing a special rule that says that one does not fire back until one gets hit. The ground level causal learning process to be described below must be able to account for this situation and determine whether the causality is indeed spurious.

III. A Ground Level Causal Learning Algorithm

Imagine an agent first arrives at a totally new environment that he has not experienced before and he observes a number of events in succession. There are many possible combinations of cause-effects between these events. A good way for an agent to begin building his causal knowledge about the environment would be to first postulate causal connections between only successive events. Later, if more evidence accrues that there could be causality between events there are intervened by other events, such as in the case shown in Fig. 1(a), the agent then retroactively corrects the earlier assumption. To do this, the agent relies on opportunistic observations, such as the non-interleaved situation in Fig. 1(a).

Fig. 2 shows an example situation using a number of familiar events, such as lightning (L), thunder (T), wind (W), vehicular sound (S), and vehicular headlight (H). These could all be experienced by an observing-agent situated at a certain location, and the agent is to work out the causalities between them. Given these events, we know that only lightning and thunder has a consistent causality and the others could happen at random times. These other events could also intervene between lightning and thunder, thus contributing "noise" to the lightning and thunder situation. The challenge for the observing-agent is to correctly recover the lightning → thunder causality.

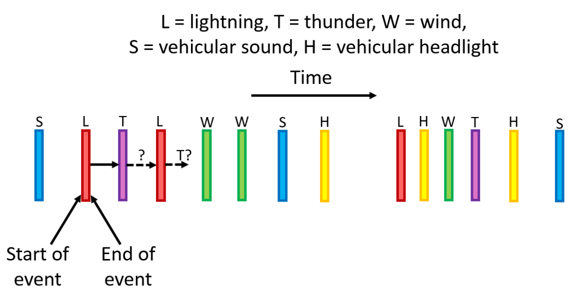


Fig. 2. Lightning, thunder, and other events.

In Fig. 2, we show that each event consists of a moment at which it starts (appears) and another moment at which it ends (disappears). The *cause* of an event could be the start or the end of a preceding event. For this paper, we simplify the situation and consider only the start of the event as the potential cause of the corresponding effect and we also consider each event as lasting over a relatively short time interval (i.e., if it is the thunder, it does not rumble over a long time).

A. *Forward and Backward Immediate Probabilities*

As mentioned above, the first step of the process is to link up successive events as cause-effect pairs. Therefore, in Fig. 2, as time progresses, these cause-effect pairs will be obtained: S→L, L→T, T→L, L→W, W→W, ….

The next step of the process is to compute what is termed the forward immediate probability (FIP) and backward immediate probability (BIP). For a given cause-effect pair, say, L→T, the FIP means the probability T is observed to follow L immediately and BIP means the probability L is observed to precede T immediately. In the example in Fig. 2, after the 4^{th} event, which is an L event, no T event is observed to follow L immediately and hence the FIP of L→T is 0.5 at this point in time. Similarly, if at some point T is observed with no immediately preceding L, then the BIP for L→T is likewise adjusted.

Even though commonsensically we know that it is lightning that causes thunder and not the other way around, the situation in Fig. 2 is such that after the 4^{th} event, which is the lightning event, the algorithm would first propose a T→L causal relationship as well because there is an L event that immediately succeeds a T event. T→L will be assigned a lower likelihood than L→T only after the consideration of other factors in Section D below.

B. *The Skipped-Over Probabilities and Consideration Threshold*

As discussed in connection with Fig. 1(a), after having established some confidence in a certain cause-effect pair based on opportunistic situations (the non-interleaved events region in Fig. 1(a)), subsequently the reasoning process, after encountering the cause, will skip over intervening "noisy" events and look for confirmation that the effect does indeed follow some time later. If instead the same cause is encountered again after skipping over many noisy events, that means the expected effect does not occur or is not observable. This reduces the probability of the cause-effect pair involved. Therefore, in the same vein as the above FIP and BIP, we define and compute forward skipped-over probability (FSP) and backward skipped-over probability (BSP).

The skipped over events are only tracked and the corresponding FSP and BSP computed if the FIP for the cause-effect pair has exceeded certain threshold. Because there could be more than one cause-effect pair that has high FIP, and among these the highest FIP value may not be 1, we use the highest FIP value as a reference and take a certain percentage of that as a threshold, called the Skipped-Over Consideration Threshold (SOCT). For those cause-effect pairs which have FIP higher than SOCT, their FSPs and BSPs are computed.

C. *The Probability of a Cause-Effect Pair*

We compute a basic probability of a cause-effect pair by taking the probability of observing an effect following a cause, minus the probability that an effect is observed without the cause preceding it. This is similar in spirit to the idea behind the contingency model of causal induction in psychology. The contingency model states that [9, 10]:

$$\Delta P_i = P(e \mid i) - P(e \mid \neg i) \quad (1)$$

where ΔP_i is the contingency between a candidate cause i and an effect e, $P(e \mid i)$ is the probability of the effect occurring when i has occurred, and $P(e \mid \neg i)$ is the probability of the effect occurring when i has not occurred. If both probabilities are the same, there is no causal connection between the cause and effect. In the same spirit, and with consideration of "skipped-over" events, we define the probability that an event EV1 is the cause of another event EV2 as:

$$Prob(Cause(EV1, EV2)) = FIP+FSP-(1-(BIP+BSP)) \quad (2)$$

where FIP, FSP, BIP, and BSP are as defined in Sections III.A and III.B above.

D. Other Factors of Causal Measure

Suppose we have the same pair of events EV1 and EV2 repeated many times, as in EV1, EV2, EV1, EV2…, it would be difficult to tell which of EV1 and EV2 is the cause and which is the effect based on the consideration so far. Ho [1] describes two situations in which it is possible to attribute one of the events as the cause and the other as the effect. In one situation, EV1 is an action emitted by an agent and EV2 is an event observed by the agent. Because the agent "knows" that EV1 is emitted by her, and she observes EV2 subsequently, she could possibly conclude that EV2 is the effect of EV1, based on the calculation of Eqn. 2 above. However, unless her internal sensory and cognitive processes tell her that she emits EV1 as a consequence of observing EV2, she knows that EV1 is not caused by EV2 and the direction of causality is only in one direction, i.e., EV1→EV2.

In another situation, both EV1 and EV2 are events observed by an observing-agent. In this case the observer-agent has no information on the internal states of the other agents that emit EV1 and EV2 and has to derive the causality based on observation alone. This could be based on an asymmetry between EV1→EV2 and EV2→EV1 that can arise under two conditions. One is due to the intervals between events. Suppose the last EV2 occurred a long time ago and EV1 is the "first" event observed, followed soon by EV2, and so on. The interval EV2→EV1 is hence much larger that the interval EV1→EV2. This results in an asymmetry between EV1→EV2 and EV2→EV1. This is the situation, say, of lightning (EV1) and thunder (EV2). (Because occasionally the interval EV1→EV2 could be larger than EV2→EV1, we consider the *average* interval over a number of event pairs.) From the point of view of effective causality [1, 5], we can say that EV1 is a better predictor of EV2, rather than vice versa.

The other measure of asymmetry is based on interval variability. The smaller the variability, the better the first event is a predictor of the second event. Therefore, a combination of interval length + interval variability (of the interval between two events) can be a measure of the predictive capability of the first event for the second.

In a special situation when there is a continuous stream of EV1 and EV2, in the form … EV1, EV2, EV1, EV2, … in which the observing-agent sees no "beginning" – i.e., no matter how far "back," there is an EV2 that precedes an EV1 and vice versa – and the interval of EV1→EV2 is the same as EV2→EV1, with both having 0 variability, then a perfectly symmetrical situation is obtained between EV1 and EV2. One can then conclude that both EV1→EV2 and EV2→EV1 are equally good. I.e., one can say that EV1 causes EV2 and EV2 causes EV1. Ho [1, 5] has established that what matters to an intelligent system is *effective* causality. In this situation, since both EV1 and EV2 can be used to reliably predict EV2 and EV1 respectively, both causal relationships are equally useful for problem solving situations.

We define the uncertainty associated with interval length, UIL, for the event pair, EV1 and EV2, as:

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

where Av(IL(EV1, EV2)) is the average interval length of all the EV1→EV2 pairs and Av(IL(∀ EVP)) is the average intervals for all event pairs. F_1 is the function:

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

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

The purpose of F_1 is to keep the value of UIL between 0 and 1, so that it is comparable with the Prob value from Eqn. 2. For $x \leq 1$, which means the average interval length of EV1 and EV2 is smaller than that of the other event pairs, UIL is 0, which means the "uncertainty" arising from the interval length value is none. For $x > 1$, the function is a logistic function which asymptotes at 1. The value of k controls the rate of change of F_1. For $k = 0.2$, it means when x is 10, which translates to the interval for the current event pair EV1, EV2 being 10 times that of the average of all event pairs, the value of $F_1(x)$ is about 0.72, which is a value close to the "worst" possible value of 1.

We define the uncertainty associated with the interval length variability, UIV, for the event pair EV1 and EV2, as:

$$UIV(EV1, EV2) =$$

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

where Dv(IL(EV1, EV2)) is the deviation of the interval length values for event pair EV1 and EV2, and Av(IL(EV1, EV2)) has the same meaning as in Eqn. 3 (there are many ways to define Dv, but a quick and dirty definition is to simply take the maximum value minus the minimum value). F_2 is the function:

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

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

Similar to the purpose of F_1 , F_2 is to keep the value of UIV between 0 and 1, so that it is comparable with the Prob value from Eqn. 2. For $x = 0$ (x is always 0 or positive in the case here), which means the interval length has 0 variability, the uncertainty associated with interval variability is hence none.

In the same vein as interval length, there could also be intervening "noise" in the form of other events (OE) that appear between the events that are in a potential cause-effect relationship that reflects the uncertainty of the causal relationship involved. We therefore use the number of intervening events between a pair of events EV1 and EV2 that are supposed to be in a causal relationship as a measure of the "uncertainty" associated with that relationship.

We define the uncertainty associated with intervening noise, UIN, for the event pair, EV1 and EV2, as:

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

where Av(OE(EV1, EV2)) is the average number of intervening events of all the EV1$\rightarrow$EV2 pairs and Av(OE($\forall$ EVP)) is the average number of intervening "other" events for all event pairs that have exceeded the SOCT as defined in Section III.B (which includes also EV1$\rightarrow$EV2). F_1 is as defined in Eqn. 4.

There is also a variability associated with the number of intervening "other" events which we term UINV and it is defined as follows:

$$UINV(EV1, EV2) =$$

$$F_2(Dv(OE(EV1, EV2)) / Av(OE(EV1, EV2)) \qquad (8)$$

where Dv(OE(EV1, EV2)) is the deviation in the number of intervening "other" events for event pair EV1 and EV2 and Av(OE(EV1, EV2)) has the same meaning as in Eqn. 7. F_2 is the same as that in Eqn. 6.

Incorporating these uncertainty measures, we define the total uncertainty, TU, for EV1 and EV2 in a causal relationship as:

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

where UIL, UIV, UIN, UINV are as defined in Eqns. 3, 5, 7, and 8 above.

Combining this total uncertainty with Eqn. 2 to arrive at a measure of the *strength* of the causal relationship between EV1 and EV2, we define Strength(Cause(EV1, EV2)) as follows:

$$Strength(Cause(EV1, EV2)) = Prob(Cause(EV1, EV2)) - w_1 * TU(Cause(EV1, EV2)) + w_2 * TU(Cause(EV2, EV1)) \qquad (10)$$

where w_1 and w_2 are the weightages of TU(Cause(EV1, EV2)) and TU(Cause(EV2, EV1)) respectively. This measure takes the basic probability of the causal relationship between EV1 and EV2 and subtract the total uncertainty associated with EV1$\rightarrow$EV2 but add the total uncertainty associated with the opposite relationship EV2$\rightarrow$EV1 to it. w_1 and w_2 are typically set at 0.5 to equalize the contributions from the various terms for the total value of Strength.

IV. Testing the Algorithm

In this section we use some simulated data to test the above ground level causal learning algorithm described in Section III.

A. Lightning and Thunder

In Fig. 3 we show a series of events happening over time. The main events of interest are lightning (L) and thunder (T). The intervals for L$\rightarrow$T tend to be closer than that for T$\rightarrow$L and tend to be of less variability. This is typical of the situation in the real world for lightning and thunder. Among the events of L and T we added "noise" in the form of three kinds of other events – wind (W), vehicular headlights (H), and vehicular sounds (S). These events occur in no definite regularity and are scattered over time randomly (though in principle there is some correlation between vehicular headlights and sounds, we are assuming the correlation is weak for an observer situated at a specific location as the headlights may not always point at her). We have also put in a situation in which L is not followed by T and another situation in which T is not preceded by L.

The Strength(Cause(EV1, EV2)) values based on Eqn. 10 are calculated for various event pairs and the results are shown in Table I (only event pairs with positive values for Strength are shown). It can be seen that L$\rightarrow$T (0.884) has the highest value, followed by T$\rightarrow$ L (0.493) and the other event pairs (S$\rightarrow$W, 0.481, H$\rightarrow$L, 0.074, etc.). The separation between T$\rightarrow$L and L$\rightarrow$T in terms of "causal strength" is not as high as we normally experience in the real world. This is primarily because there is one feature in the real world that is not captured in Fig. 3, which is that often after the last set of lightning-thunder, there could be no lightning-thunder for many days, and this would increase the T$\rightarrow$L average interval as well as interval variability by a large amount, leading to a much lower T$\rightarrow$L Strength.

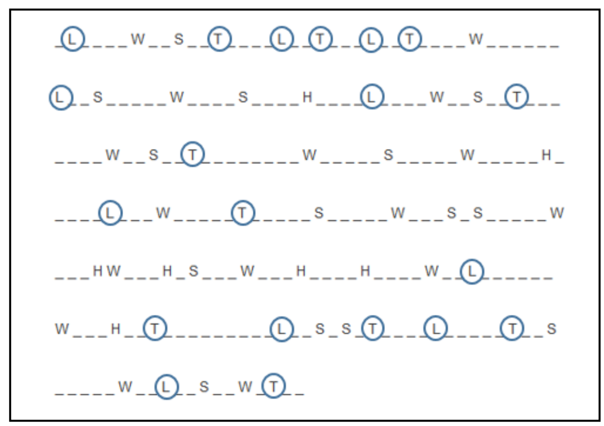


Fig. 3. Test data: lightning (L), thunder (T), and other events (W = wind, H = vehicular headlight, S = vehicular sound). L and T are circled in blue for ease of identification. Events begin from top left corner.

Similarly, causalities such as S$\rightarrow$W or W$\rightarrow$H do not exist in the real world but they obtain some significant values here. We can think of the situation as due to the unusually large number of sound and wind events during this "lightning and thunder" event period. An observer-agent new to the world who observes this sequence of events the first time would also see this causality between S and W.

TABLE I. Cause-Effect Strengths of Fig. 3 Events

Cause$\rightarrow$Effect	***Strength***
L$\rightarrow$T[a]	0.884
T$\rightarrow$L	0.493
S$\rightarrow$W	0.481
W$\rightarrow$S	0.403

Cause→Effect	*Strength*
W→H	0.311
L→W	0.295
H→L	0.074
W→L	0.047
S→T	0.022

a. Only event pairs with positive Strength values are listed

B. Lightning and Thunder with Firing and Pain

In this section we test the ability of the algorithm to identify more than one strongly linked causal relationship. In Fig. 4 we added the event Firing (F) followed consistently by the event Pain (P) (like those in Fig. 1(a)) among a similar set of other events of Fig. 3, including lightning (L) and thunder (T). A number of F and P pairs of events are intervened by other "noisy" events, including L.

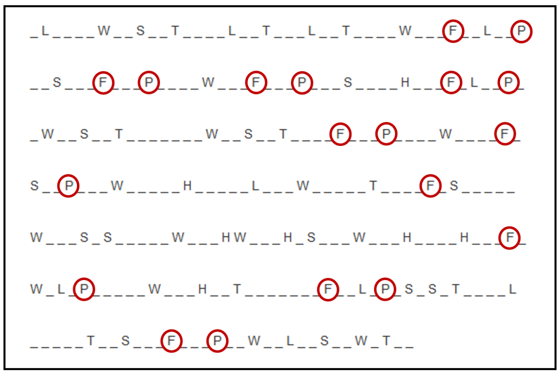


Fig. 4. Test data: lightning (L), thunder (T), Firing (F), Pain (P), and other events (W = wind, H = vehicular headlight, S = vehicular sound). F and P are circled in red for ease of identification. Events begin from top left corner.

Table II shows the results of applying Eqn. 10 to the data of Fig. 4. It can be seen in Table II that both L→T and F→P obtain the highest Strength scores. Therefore, both highly correlated event pairs have been successfully recovered from the seemingly noisy event data stream.

TABLE II. CAUSE-EFFECT STRENGTHS OF FIG. 4 EVENTS

Cause-Effect	*Strength*
L→T[b]	0.857
F→P	0.627
T→L	0.521
W→H	0.351
W→S	0.336
L→W	0.157
F→L	0.073
W→F	0.046
P→S	0.027
P→W	0.007

b. Only event pairs with positive Strength values are listed

V. CONCLUSION

In this paper we developed a ground level causal learning algorithm to uncover (effective) causalities between events in the presence of noise. We developed a number of measures to gauge whether two events have a good causal relationship and these are added up in a total measure of Strength of causal connection (Eqn. 10). The test results show that the algorithm is able to satisfactorily identify pairs of events that are strongly correlated temporally and are hence causally connected based on the Strength measure. Future work calls for the application of the method to real-world situations.

ACKNOWLEDGMENT

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

REFERENCES

[1] S.-B. Ho, Principles of Noology: Toward a Theory and Science of Intelligence. Switzerland: Springer, 2016.

[2] Pearl, Causality: Models, Reasoning, and Inference, 2nd Edition, Cambridge: Cambridge University Press, 2009.

[3] P. W. Cheng, "From covariation to causation: A causal power theory," Psychological Review, vol. 104, no. 2, pp. 367-405, 1997.

[4] S. G. B. Johnson and F. C. Keil, "Causal inference and the hierarchical structure of experience," Journal of Experimental Psychology: General, vol. 143, no. 6, pp. 2223-2241, 2014.

[5] S.-B. Ho, "On effective causal learning," in Proceedings of the 7th International Conference on Artificial General Intelligence, pp. 43-52, Berlin: Springer Verlag, 2014.

[6] A. Agresti and C. Franklin, Statistics: The Art and Science of Learning from Data, 3rd ed. Boston: Pearson Education, Inc., 2007.

[7] D. S. Moore, G. P. McCabe, and B. A. Craig, Introduction to the Practice of Statistics, 6th ed. New York: W. H. Freeman, 2009.

[8] S.-B. Ho, "A grand challenge for computational intelligence" in Proceedings of the IEEE Symposium Series on Computational Intelligence for Intelligent Agents, pp. 44-53, Piscataway, New Jersey: IEEE Press, 2013.

[9] H. Jenkins and W. Ward, "Judgment of contingency between responses and outcomes," Psychological Monographs, vol. 7, pp. 1-17, 1965.

[10] R. A. Rescorla, "Probability of shock in the presence and absence of CS in fear conditioning," Journal of Comparative and Physiological Psychology, vol. 66, pp. 1-5, 1968.