Graph Weaknesses In Commonsense Causal Representation

LAWRENCE J. MAZLACK
Applied Computational Intelligence Laboratory
University of Cincinnati
Cincinnati, Ohio 45221
mazlack@uc.edu

Abstract: - Causal reasoning occupies a central position in human reasoning. In order to algorithmically consider causal relations, the relations must be placed into a representation that supports manipulation. The most widespread causal representation in current usage is directed acyclic graphs. However, they are severely limited in what portion of the every day world they can represent. Some of the required Markov conditions do not fit with commonsense reasoning. More importantly, cycles must be represented and they cannot be represented in acyclic graphs. Additionally, shifts in grain size are overly limited. Commonsense understanding deals with imprecision, uncertainty and imperfect knowledge. An algorithmic way of handling and representing causal imprecision that includes cycles is needed.

Key-Words: - causality, representation, commonsense reasoning, directed acyclic graphs, complexes, imprecision

1 Introduction
Causal reasoning occupies a central position in human reasoning. It has a core position in human decision-making. Considerable effort has been spent examining causation. Whether causality can be precisely defined or can be recognized at all is theoretically uncertain. At the same time, people operate on the commonsense belief that causality exists. This paper informally discusses some of the issues with a focus on the suitability of directed graph representation.

Of particular interest to this paper are areas where the analysis is observational (non-experimental). Data mining is of interest; one product of data mining is association rules.

Customers who
buy beer and sausage
also tend to buy hamburger
with (confidence = 0.7)
in (support = 0.2)

Customers who buy strawberries
also tend to buy whipped cream
with (confidence = 0.8)
in (support = 0.15)

Fig. 1. Association rules.

At first glance, association rules seem to imply a causal or cause-effect relationship. That is:

A customer’s purchase of both sausage and beer causes the customer to also buy hamburger.

But, all that is discovered is the existence of a statistical relationship between the items. They have a degree of joint occurrence. The nature of the relationship is not understood. It is not known whether the presence of an item or set of items causes the presence of another item or set of items; or the converse, or some other phenomenon causes them to occur together. The computational need of data mining is to recognize what relationships are causal and which are not.

Purely accidental relationships do not have the same decision value, as do causal relationships. For example,

IF it is true that buying strawberries somehow causes someone to buy whipped cream.

THEN: A merchant might profitably put strawberries on sale

AND at the same time: Increase the price of whipped cream to compensate for strawberries sale price

On the other hand, knowing that

Bread and milk are often purchased together.

may not be useful information as the purchase of one does not influence the purchase of the other; both are commonly purchased on every store visit, generally independently.

Association rules do not necessarily describe causality; i.e., the degree that one thing is caused to happen by another thing. The confidence measure is simply an estimate of conditional probability. Support indicates how often the joint occurrence happens (the joint probability over the entire data set). The joint occurrence count is symmetric; that is, it does not matter what we count first. Also, the strength of any causal dependency may be
very different from that of a paired association value. In all cases

\[ confidence \geq \text{causal dependence} \]

All that can be said is that associations describe the strength of joint co-occurrences.

Association rules can be used is to aid in making retail marketing decisions. However, naïve use of association rules may lead to errors. Errors might occur; either if causality is assumed where there is no causality; or if the direction of the causal relationship is wrong [1] [2]. False assumptions can lead to unfounded decisions. For example, if

A study of our customers shows that 94% are sick.

- Is it the following possible causal rule?
  - Our customers are sick, so they buy from us.
- Or, is it the following possible complementary causal rule?
  - If people use our products, they are likely to become sick.

From a decision making viewpoint, it is not enough to know that

People both buy our products and are sick.

This is the sole causal element; the other nested constraints and laws the complex is subject to.

When events happen, there are usually other related events. The entire collection of events can be called a complex. A “mechanism” [5] or a “causal complex” [6] [7] is a collection of events whose occurrence or non-occurrence results in a consequent event happening. Hobbs [6] suggests that human causal reasoning making use of a causal complex does not require precise, complete knowledge of the complex.

Each complex, taken as a whole, can be considered to be a granule if the grain size is increased. Larger complexes can be decomposed into smaller complexes; going from large grained to small grained. For example, when describing starting an automobile. A large-grained to small-grained, nested causal view could start with

When an automobile's ignition switch is turned on, this causes the engine to start.

But, it would not happen if a large system of other nested conditions were not in place.

There has to be available fuel. The battery has to be good. The switch has to be connected to the battery so electricity can flow through it. The wiring has to connect the switch to the starter and ignition system (spark plugs, etc.). The engine has to be in good working order; and so forth.

Turning the ignition switch on is one action in a complex of conditions required to start the engine. The largest grained view is: turning on the switch is the sole causal element; the complex of other elements represents the finer grains. These elements in turn could be broken down into still finer grains; for example, “available fuel” could be broken down into:

- fuel in tank, working fuel pump, intact fuel lines, etc.

Sometimes, it is enough to know what happens at a large grained level; at other times it is necessary to know the finer grained result. For example, if

Bill believes that turning the ignition key of his automobile causes the automobile to start.

It is enough if

Bill engages an automobile mechanic when his automobile does not start when he turns the key on.

As the automobile mechanic knows a finer grained view of an automobile’s causal complex than does Robin.

Nested granularity may be applied to causal complexes. A complex may be several larger grained elements. In turn, each of the larger grained elements may be a complex of more fine-grained elements. In turn, these elements may be made up still finer grained
elements. In general, people are more successful in applying commonsense reasoning to a few large grain sized events than the many fine-grained elements that might make up a complex. A useful causal representation must be able to both represent complexes and shift between grain sizes.

3 Precisely Defining Causal Relationships

Coming to a precise description of what is meant by causality is difficult. There are multiple and sometimes conflicting definitions. Regardless, we do have a commonsense belief that there are causal relationships.

A common, simplistic conception [8] is that causality depends on one-way, time ordering. In contrast, Simon [9] [10] provides an analysis of causality that does not rely on time order. For a more extensive definition of various aspects of causality definition, see Mazlack [11].

Perhaps, complete knowledge of all possible situational factors might lead to a crisp description of whether an effect will occur. However, it is unlikely that the identity all possible elements can or may come to be known. It is also unlikely that it may be possible to fully know the certainty all of the elements involved.

3.1. Positive Causation

Naïve commonsense understanding of causation exclusively focuses on simple positive causation. The idea is that one thing causes another to happen; for example,

When a glass is pushed off a table and breaks on the floor
it might be said that
Being pushed from the table caused the glass to break.

Although,
Being pushed from a table is not a certain cause of breakage; sometimes the glass bounces and no breakage occurs; or, someone catches the glass before it hits the floor.

Counterfactually and more weakly,
Not falling to the floor prevents breakage.

A counter factual statement is weaker as a causal statement as other positive causative factors can still come into play; for example, sometimes
A glass breaks when an errant object hits it, even though it does not fall from the table.

Positive causal relationships can be described as: if $\alpha$ then $\beta$ (or, $\alpha \rightarrow \beta$). For example:

\[\text{When an automobile driver fails to stop at a red light and there is an accident it can be said that the failure to stop was the accident's cause.}\]

3.2 Negation

Negation can come into play both because of the absence of a positive factor or because of a factor that prevents something from happening. Negating the causal factor does not mean that the effect does not happen; sometimes effects can be overdetermined. For example:

An automobile that did not fail to stop at a red light can still be involved in an accident; another car can hit it because the other car's brakes failed.

In general, negative statements are weaker than positive statements. Negative statements can become overextended. It cannot be said that $\neg\alpha \rightarrow \neg\beta$, for example:

Failing to stop at a red light is not a certain cause of an accident occurring; sometimes no accident occurs whether or not there was a stop.

besides overextension, effects can be overdetermined; that is: more than one item can cause an effect. If so, eliminating one cause does not necessarily eliminate the effect. For example, the rule:

If a person drinks wine, they may become inebriated.

cannot be simply negated to

If a person does not drink wine, they will not become inebriated.

In this case:

A person may also drink beer or whiskey to excess and become inebriated.

Events that do not happen can similarly be overdetermined. For example, for Ortiz [12] states that it is not true that

His closing the barn door caused the horse not to escape

because the horse might not have attempted to escape even if the door was open. Therefore, a false counterfactual is:

If he had not closed the barn door, the horse would have escaped.

Similarly, for example, the rule

If a person smokes, they will get cancer.
cannot be simply negated to

If a person does not smoke, they will not get cancer.

Again, effects can be overdetermined. In this case,

People who do not smoke may still get cancer from other causes.

4 Causality Recognition & Representation

Various causality descriptions and discovery tools have been suggested. It may eventually turn out that different subject domains may have different methodological preferences.

Hobbs [6] uses first order logic to describe causal complexes. Pearl [13] develops probabilistic causal networks of directed graphs (DAGs). The causal complexes explicitly considered by Hobbs and Pearl have a required structure that may be overly restrictive for commonsense causal understanding, namely:

- If all of the events in the causal complex appropriately happen, then the effect will occur
- There is nothing in the causal complex that is irrelevant to the effect

These requirements are probably too precise and extensive to be realized in a commonsense world. Sometimes, only some of the events need to happen. For example,

Someone may be able to save more money:
- If their taxes are lowered or
- If they earn more money

Either even may lead to greater savings. However,

Neither may result in increased savings if they also have to pay a large divorce settlement.

So, if all of the events happen, the effect may happen. If some of the events happen, the effect may happen. We rarely know whether all of the events are in a complex are necessary. For example,

A man may want to attract the attention of a woman. He may do a large number of things (e.g., hair, clothes, learn to dance, etc.). If he does attract the woman, he may never know which things were relevant and which were not.

5 Directed Graphs And Causality

The idea of “positive” causation ($\alpha \rightarrow \beta$) is at the core of commonsense causal reasoning. Often a positive causal relationship is represented as a network of nodes and branches [14].

Various graph based Bayesian based methods have been suggested to describe causality. Probably the best known is the class of methods based on Directed Acyclic Graphs (DAGs). The most fully developed approach is Pearl [13]. Silverstein [15] [16] followed a similar approach.

From the commonsense causal reasoning view, the various acyclic directed graph methods have similar liabilities, specifically:

5.1 Liability: Cyclic needs that cannot be represented in a DAG.

Causal relationships cannot be cyclic in a DAG, either directly or indirectly (through another attribute). This is at variance with our commonsense understanding of the world. Within cyclic dependencies, there are variants.

5.1.1 Liability: Representing cycles with time lag: feedback There are many commonsense examples where cycles are needed.

Fig. 4. Positive feedback cycle: Robin tells Kim that I love you. Then, Kim tells Robin I love you. Then, Robin tells Robin I love you more than before. Then, Kim ... and so forth and so on. The cyclic reinforcement could be substantial.

Fig. 5. Cyclic relationship that can be collapsed if process knowledge loss is acceptable.

Fig. 6. Cyclic relationship that cannot be collapsed
In some cases, some cycles can be reasonably collapsed, some cannot be. In Fig. 5, perhaps the right hand cycle can be collapsed without a loss of significant meaning. However, in the Fig. 6, information would be lost.

5.1.2 Liability: Representing: concurrent cycles. A form of a cycle is concurrent joint mutual dependency with no time lag. Mutual dependencies are possible; i.e., \( \alpha \rightarrow \beta \) as well as \( \beta \rightarrow \alpha \). It seems to be possible that they do so with different strengths with equal strengths being a special case. The dependencies can be described as shown in the following figure where \( S_{ij} \) represents the strength of the causal dependency from \( i \) to \( j \).

![Cyclic relationship: Mutual dependency.](image)

Fig. 7. Cyclic relationship: Mutual dependency.

There are two variations: differing causal strengths for the same activity; and, different causal strengths for symmetric activities occurring at different times.

5.1.2.1 Different causal strengths for the same activity, simultaneously occurring

Some argue that causality should be completely asymmetric and if it appears that items have mutual influences it is because there is another cause that causes both. A problem with this is that it can lead to eventual regression to a first cause. Whether this is true or not, it is not useful for commonsense representation. In contrast, Simon [5] and Shoham [17] identify cases where causality is simultaneous.

It is also our commonsense experience. For example, in the preceding figure, \( \alpha \) could be short men and \( \beta \) could be tall women. If \( S_{\alpha \beta} \) meant the strength of desire for a social meeting that was caused in short men by the sight of tall women, it might be that \( S_{\alpha \beta} > S_{\beta \alpha} \).

5.1.2.2 Different causal strengths for symmetric activities, occurring at different sequential times

It would seem that if there were causal relationships in market basket data, there would often be imbalanced dependencies. For example, if

A customer first buys strawberries, there may be a reasonably good chance that they will then buy whipped cream.

Conversely, if

They first buys whipped cream, the subsequent purchase of strawberries may be less likely.

5.2 Liability: Meeting Markov Conditions

DAGs require that several Markov conditions be met. These requirements sometimes do not fit well with commonsense understanding of the world.

5.2.1 Markov Liability: Stationary Condition: Probabilities are time independent

This does not correspond to our commonsense understanding of the world. If one event is dependent on two other causal events, if one causing event happens much earlier (or later) than the other causing event, there may well be a different result.

![Case where differing times in causal events affects probability of causal result.](image)

Fig. 8. Case where differing times in causal events affects probability of causal result.

5.2.2 Markov Liability: Memoryless States

The Memoryless Markov Condition is defined as: Let \( A \) be a node in a causal Bayesian network, and let \( B \) be any node that is not a descendant of \( A \) in the network. Then the Markov (Markoff) condition holds if \( A \) and \( B \) are independent, conditioned on the parents of \( A \). The intuition of this condition is: If \( A \) and \( B \) are dependent, then \( B \) must either be (a possibly indirect) cause of \( A \) or (possibly indirectly) caused by \( A \). In the second case, \( B \) is a descendant of \( A \), while in the first \( B \) is an ancestor of \( A \) and has no effect on \( A \) once \( A \)'s immediate parents are fixed.

![“Memoryless” Markov condition holds.](image)

Fig. 9. “Memoryless” Markov condition holds.
This makes sense in Fig. 9. However, not all of our commonsense perceptions of causality work this way. Often, we believe that history matters as in Fig. 10.

6 Conclusions

Causal reasoning occupies a central position in human reasoning.

Commonsense reasoning recognizes granularization and that objects may be made up out of granules. Knowledge of at least some causal effects is imprecise. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether an effect will occur. However, it is unlikely that all possible factors can be known. A causal model must accommodate shifts in grain size as well as imprecision and incompleteness.

There are several needs of a causal model. Some are:
- Represent imprecision: DAGs OK
- Accommodate changes in grain size: DAGs not OK as not all cycles can be collapsed
- Describe complexes: DAGs OK
- Avoid over determination and over extension: Graphs in general, including DAGs, not OK
- Support cyclic models of all kinds: DAGs not OK as cannot represent any kind of cycles
- Be time varying: DAGs not OK
- Not be restricted by Markov conditions: DAGs not OK, they must meet certain Markov conditions (stationary, Memoryless)
- Handle incompleteness: DAGs, not OK; other models such as fuzzy model are better

In order to algorithmically consider causal relations, the elements must be placed into a representation that supports manipulation. The most widespread causal representation is directed acyclic graphs (DAGs). However, DAGs are severely limited in what portion of the common sense world they can represent; other representation methods must be used.

References