Causality Recognition For Data Mining In An Inherently Ill Defined World

Lawrence J. Mazlack
Applied Artificial Intelligence Laboratory
University of Cincinnati
Cincinnati, OH 45221-0030
mazlack@uc.edu

Abstract

Commonsense causal reasoning occupies a central position in human reasoning. It plays an essential role in both informal and formal human decision-making. Causality itself as well as human understanding of causality is imprecise, sometimes necessarily so. Our common sense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. A difficulty is striking a good balance between precise formalism and commonsense imprecise reality. Clearly, an algorithmic method of handling imprecision is needed. Today, data mining holds the promise of extracting unsuspected information from very large databases. In many ways, the interest is the promise (or illusion) of causal, or at least, predictive relationships. However, the most common data mining rule forms only calculate a joint occurrence frequency; they do not express a causal relationship. Without understanding the underlying causality, a naïve use of data mining rules can lead to undesirable actions.

1. Introduction

Commonsense causal reasoning occupies a central position in human reasoning. It plays an essential role in human decision-making. Considerable effort has been spent examining causation. Philosophers, mathematicians, computer scientists, cognitive scientists, psychologists, and others have formally explored questions of causation beginning at least three thousand years ago with the Greeks.

Whether causality can be recognized at all has long been a theoretical speculation of scientists and philosophers. At the same time, in our daily lives, we operate on the commonsense belief that causality exists.

Causal relationships exist in the commonsense world. If an automobile 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. However, conversely, failing to stop at a red light is not a certain cause of a fatal accident; sometimes no accident of any kind occurs. So, it can be said that knowledge of some causal effects is imprecise. Perhaps, complete knowledge of all possible factors might lead to a crisp description of whether a causal effect will occur. However, in our commonsense world, it is unlikely that all possible factors can be known. What is needed is a method to model imprecise causal models.

Another way to think of causal relationships is counterfactually. For example, if a driver dies in an accident, it might be said that had the accident *not* occurred; they would still be alive.

Our common sense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. This is also the case of our scientific knowledge of the world. Clearly, we need an algorithmic way of handling imprecision if we are to computationally handle causality. Models are needed to algorithmically consider causes. These models may be symbolic or graphic. A difficulty is striking a good balance between precise formalism and commonsense imprecise reality.

1.1 Data Mining, Introduction

Data mining is an advanced tool for managing large masses of data. It analyzes data previously collected. It is *secondary* analysis. Secondary analysis precludes the possibility of experimentally varying the data to identify causal relationships.

There are several different data mining products. The most common are *conditional rules* or *association rules*. Conditional rules are most often drawn from induced trees while association rules are most often learned from tabular data. Of these, the most common data mining product is association rules; for example:

• ***Conditional rule:***
IF Age < 20
THEN Income < $10,000
with {belief = 0.8}

• ***Association rule:***
Customers who
buy beer and sausage
also tend to buy mustard
with {confidence = 0.8}
in {support = 0.15}

At first glance, these 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 mustard.* In fact, all that is discovered is the *existence* of a statistical relationship between the items. The *nature* of the relationship is not specified. We do not know whether the presence of an item or sets of items causes the presence of another item or set of items; or the converse, or some other phenomenon causes them to occur together.

When typically developed, association rules do not *necessarily* describe causality. Also, the strength of causal dependency may be very different from a respective association value. All that can be said is that associations describe the strength of joint co-occurrences. Sometimes, the relationship might be causal; for example, if someone eats salty peanuts and then drinks beer, there is probably a causal relationship. On the other hand, it is unlikely that a crowing rooster causes the sun to rise.

1.2 Naive Association Rules Can Lead To Bad Decisions

One of the reasons why association rules are used is to aid in making retail decisions. However, simple association rules may lead to errors. It is common for a food store to put one item on sale and then to raise the price of another item whose purchase is assumed to be associated. This may work if the items are truly associated; but it is problematic if association rules are blindly followed [Silverstein, 1998].

Example: At a particular store, a customer buys:

- *hamburger 33%* of the time
- *hot dogs* 33% of the time
- both *hamburger* and *hot dogs* 33% of the time
- *sauerkraut*[*] only if *hot dogs* are also purchased

This would produce the transaction matrix:

	hamburger	hot dog	sauerkraut
t_1	1	1	1
t_2	1	0	0
t_3	0	1	1

This would lead to the associations:

- (hamburger, hot dog) = 0.5
- (hamburger, sauerkraut) = 0.5
- (hot dog, sauerkraut) = 1.0

[*] Sauerkraut is a form of pickled cabbage. Some people greatly enjoy using sauerkraut as a garnish with sausages. However, it is rarely consumed as a garnish with hamburger. For more about sauerkraut, see: *http://www.sauerkraut.com/*

If the merchant:

- Reduced price of hamburger (as a sale item)
- Raised price of sauerkraut to compensate (as the rule *hamburger* $\Rightarrow$ *sauerkraut* has a high confidence.
- The offset pricing compensation would not work as the sales of sauer-kraut would not increase with the sales of hamburger. Most likely, the sales of hot dogs (and consequently, sauer-kraut) would likely decrease as buyers would substitute hamburger for hot dogs.

1.3 False Causality

Complicating causal recognition are the many cases of false causal recognition. For example, a coach may win a game when wearing a particular pair of socks, then always wear the same socks to games. More interesting, is the occasional false causality between music and motion. For example, Lillian Schwartz developed a series of computer generated images, sequenced them, and attached a sound track (usually Mozart). While there were some connections between one image and the next, the music was not scored to the images; however, a person viewing the assemblage viewing them, the music appeared to be connected. All of the connections were observer supplied.

An example of non-computer illusionary causality is the choreography of Merce Cunningham. To him, his work is non-representational and without intellectual meaning. He often worked with John Cage, a randomist composer. Cunningham would rehearse his dancers, Cage would create the music; only at the time of the performance would music and motion come together. However, the audience usually conceived of a causal connection between music and motion and saw structure in both.

1.4 Recognizing Causality Basics

A common approach to recognizing causal relationships is by manipulating variables by experimentation. How to accomplish causal discouvery in purely observational data is not solved. (Observational data is the most likely to be available for data mining analysis.) Algorithms for discouvery in observational data often use correlation and probabilistic independence. If two variables are statistically independent, it can be asserted that they are not causally related. The reverse is not necessarily true.

Real world events are often affected by a large number of potential factors. For example, with plant growth, many factors such as temperature, chemicals in the soil, types of creatures present, etc., can all affect plant growth. What is unknown is what causal factors will or will not be present in the data; and, how many of the underlying causal relationships can be discouvered among observational data.

Some define cause-effect relationships as: When α occurs, β always occurs. This is inconsistent with our commonsense understanding of causality. A simple environment example: When a hammer hits a bottle, the bottle *usually* breaks. A more complex environment example: When a plant receives water, it *usually* grows.

An important part of data mining is understanding whether there is a relationship between data items. Sometimes, data items may occur in pairs but may not have a deterministic relationship; for example, a grocery store shopper may buy both bread and milk at the same time. Most of the time, the milk purchase is not caused by the bread purchase; nor is the bread purchase caused by the milk purchase.

Alternatively, if someone buys strawberries, this may causally affect the purchase of whipped cream. *Some* people who buy strawberries want whipped cream with them; of these, the desire for the whipped cream varies. So, we have a conditional primary effect (whipped cream purchase) modified by a secondary effect (desire). How to represent all of this is open.

A largely unexplored aspect of mined rules is how to determine when one event causes another. Given that α and β are variables and there appears to be a statistical covariability between α and β, is this covariability a causal relation? More generally, when is any pair relationship causal? Differentiation between covariability and causality is difficult.

Some problems with discouvering causality include:

- Adequately defining a causal relation
- Representing possible causal relations
- Computing causal strengths
- Missing attributes that have a causal effect
- Distinguishing between association and causal values
- Inferring causes and effects from the representation.

Beyond data mining, causality is a fundamentally interesting area for workers in intelligent machine based systems. It is an area where interest waxes and wanes; in part because of definitional and complexity difficulties. The decline in computational interest in cognitive science also plays a part. Activities in both philosophy and psychology [Glymour, 2001] overlap and illuminate computationally focused work. Often, the work in psychology is more interested in how people *perceive* causality as opposed to whether causality actually exists. Work in psychology and linguistics [Lakoff, 1990] [Mazlack, 1987] show that categories are often linked to causal descriptions. For the most part, work in intelligent computer systems has been relatively uninterested in grounding based on human perceptions of categories and causality. This paper is concerned with developing commonsense representations that are compatible in several domains.

2. Causality

Centuries ago, in their quest to unravel the future, mystics aspired to decipher the cries of birds, the patterns of the stars and the garbled utterances of oracles. Kings and generals would offer precious rewards for the information soothsayers furnished. Today, though predictive methods are different from those of the ancient world, the knowledge that dependency recognition attempts to provide is highly valued. From weather reports to stock market prediction, and from medical prognoses to social forecasting, superior insights about the shape of things to come are prized [Halpern, 2000].

Democritus, the Greek philosopher, once said: “Everything existing in the universe is the fruit of chance and necessity.” This seems self-evident. Both randomness and causation are in the world. Democritus used a poppy example. Whether the poppy seed lands on fertile soil or on a barren rock is chance. If it takes root, however, it will grow into a poppy, not a geranium or a Siberian Husky [Lederman, 1993].

Beyond computational complexity and holistic knowledge issues, there appear to be inherent limits on whether causality can be determined. Among them are:

- *Quantum Physics:* In particular, Heisenberg’s uncertainty principle
- Knowledge of the world might never be complete because we, as observers, are integral parts of what we observe
- *Gödel’s Theorem:* Which showed in any logical formulation of arithmetic that there would always be statements whose validity was indeterminate. This strongly suggests that there will always be inherently unpredictable aspects of the future.
- *Turing Halting Problem:* Turning (as well as Church) showed that any problem solvable by a step-by-step procedure could be solved using a Turing machine. However, there are many routines where you cannot ascertain if the program will take a finite, or an infinite number of steps. Thus, there is a curtain between what can and cannot be known mathematically.
- *Chaos Theory:* Chaotic systems appear to be deterministic; but are computationally irreducible. If nature is chaotic at its core, it might be fully deterministic, yet wholly unpredictable [Halpern, 2000, 139].

- *Space-Time:* The malleability of Einstein's space time that has the effect that what is "now" and "later" is local to a particular observer; another observer may have contradictory views.
- *Arithmetic Indeterminism:* Arithmetic itself has random aspects that introduce uncertainty as to whether equations may be solvable. Chatin [1987, 1990] discovered that Diophantine equations may or may not have solutions, depending on the parameters chosen to form them. Whether a parameter leads to a solvable equation appears to be random. (Diophantine equations represent well-defined problems, emblematic of simple arithmetic procedures.)

Given determinism's potential uncertainty and imprecision, we might throw up out hands in despair. It may well be that a precise and complete knowledge of causal events is uncertain. On the other hand, we have a commonsense belief that causal effects exist in the real world. If we can develop models tolerant of imprecision, it would be useful. Perhaps, the tools found in soft computing may be useful.

3. Problems With Using Probability

There has been significant work in using various forms of Bayesian networks for causal discovery. A *Bayesian network* is a combination of a probability distribution and a structural model that is a directed acyclic graph in which the nodes represent the variables (attributes) and the edges (arcs) represent probabilistic dependence. A *causal Bayesian network* is a Bayesian network where the predecessors of a node are interpreted as directly causing the variable associated with a node. However, Bayesian networks can be computationally expensive. Inferring *complete* causal Bayesian networks is essentially impossible in large scale data mining with thousands of variables.

Restricted algorithms [Cooper, 1997] have been suggested that might be useful for causal discovery in market basket data. However, the restrictions on the data and the assumptions made about the relationships are overly limiting. The restrictions are:

- Discrete or continuous data must be reduced to Boolean values
- There is no missing data
- Causal relationships are not cyclic, either directly or indirectly (through another attribute)

4. Epilogue

Causality occupies a central position in human commonsense reasoning. In particular, it plays an essential role in common sense human decision-making by providing a basis for choosing an action that is likely to lead to a desired result. In our daily lives, we make the commonsense observation that causality exists. Carrying this commonsense observation further, the concern is how to computationally recognize a causal relationship.

Data mining holds the promise of extracting unsuspected information from very large databases. Methods have been developed to build rules. In many ways, the interest in rules is that they offer the promise (or illusion) of causal, or at least, predictive relationships. However, the most common form of data mining rules (association) only calculate a joint occurrence frequency, not a causal strength. A fundamental question is determining whether or not recognizing an association can lead to recognizing a causal relationship.

An interesting question how to determine when causality can be said to be stronger or weaker. Either in the case where the causal strength may be different in two independent relationships; or, where in the case where two items each have a causal relationship on the other.

Causality is a central concept in many branches of science and philosophy. In a way, the term "causality" is like "truth" -- a word with many meanings and facets. Some of the definitions are extremely precise. Some of them involve a style of reasoning best be supported by fuzzy logic.

Defining and representing causal and potentially causal relationships is necessary to applying algorithmic methods. A graph consisting of a collection of simple directed edges will most likely not offer a sufficiently rich representation. Representations that embrace some aspects of imprecision are necessary.

A deep question is when anything can be said to cause anything else. And if it does, what is the nature of the causality? There is a strong motivation to attempt causality discouvery in association rules. The research concern is how to best approach the recognition of causality or non-causality in association rules. Or, if there is to recognize causality as long as association rules are the result of secondary analysis?

References

G. Chatin [1987] **Algorithmic Information Theory**, Cambridge University Press, Cambridge, United Kingdom

G. Chatin [1990] "A Random Walk In Arithmetic," New Scientist 125, n 1709 (March, 1990), 44-66

G. Cooper [1997] "A Simple Constraint-Based Algorithm For Efficiently Mining Observational Databases For Causal Relationships," Data Mining and Knowledge Discovery, v 1, n 2, 203-224

C. Glymour [2001] **The Mind's Arrows, Bayes Nets And Graphical Causal Models In Psychology**, MIT Press, Cambridge, Massachusetts

P. Halpern [2000] **The Pursuit Of Destiny**, Perseus, Cambridge, Massachusetts

G. Lakoff [1990] **Women, Fire, And Dangerous Things: What Categories Reveal About The Mind**, University of Chicago Press

L. Lederman, D. Teresi [1993] **The God Particle: If the Universe Is the Answer, What Is the Question?** Delta, New York

L. Mazlack [1987] "Machine Conceptualization Categories," *Proceedings 1987 IEEE Conference on Systems, Man, and Cybernetics*

C. Silverstein, S. Brin, R. Motwani, J. Ullman [1998] "Scalable Techniques for Mining Causal Structures," Proceedings. 1998 International Conference Very Large Data Bases, New York, NY, August 1998, 594--605,

L. Zadeh [2000] "Abstract Of A Lecture Presented At The Rolf Nevanilinna Colloquium, University of Helsinki," reported to: *Fuzzy Distribution List*, fuzzy-mail@dbai.tuwien.ac.at, August 24, 2000