

Causal Modeling Approximations In The Medical Domain

Lawrence J. Mazlack

Applied Computational Intelligence Laboratory
University of Cincinnati
mazlack@uc.edu

Abstract---Studies in the health sciences often seek to discover cause-effect relationships among observed variables of interest, for example: treatments, exposures, preconditions, and outcomes. Consequently, causal modeling and causal discovery are central to medical science. In order to algorithmically consider causal relations, the relations must be placed into a representation that supports manipulation and discovery. Knowledge of at least some causal effects is inherently imprecise or approximate. The most widespread causal representation is directed acyclic graphs (DAGs). However, DAGs are limited in what they can represent. Another graph methodology, fuzzy cognitive maps (FCMs) hold promise as a model that overcomes some of the difficulties found in other approaches. This paper considers causality and suggests fuzzy cognitive maps as a useful causal representation methodology.

Keywords---causal, modeling, cognitive map

I. INTRODUCTION

Causal knowledge makes up much of what we know and want to know in science [1]. Causal modeling and causal discovery are central to medical science as recognition of cause-effect can lead to robust treatments. Causality is a *relation*, relating *causes* to *effects* [2]. Experimental studies, such as randomized controlled trials, often provide the trust worthiest results for establishing causal relationships. However, human experimental studies, while potentially informative, may not be: safe, ethical, logistically feasible, or financially worthwhile. Alternatively, observational data is often available in the form of medical records; this data can be analyzed to attempt to recognize causal relationships. Sometimes, there is insufficient data to draw precise results; or, the available data is imprecise, incomplete, or contradictory; this leads to the need to accommodate imprecision in a casual model.

The target of many studies in the health sciences is the discovery of cause-effect relationships among recorded variables of interest, for example: treatments, exposures, preconditions, and outcomes [3]. Yet, until recently, the dominant methodology in health science research has been statistical analysis. Statistical analysis of recorded data can provide information about correlations (or associations), but not necessarily information about causal relationships. For example, the correlation between a patient's *coughing* and *fever* may be high. But, this does not necessarily indicate a causal relationship; i.e., someone's *cough* is *caused* by their *fever* (or the other way around). A high correlation does not indicate causality. All that a high correlation tells us is that, for some reason, certain items often occur at the same time.

Causal relationships can be described as having varying strengths; i.e., we can say that they are *always true*, *often*, *sometimes true*, *rarely true*, etc. Zadeh [4] calls approximating natural language terms: *linguistic variables*. Linguistic imprecision can be best manipulated and represented using *fuzzy arithmetic* which is based on *fuzzy sets* [5].

Statistical analysis of observational data can provide information about correlations, but not necessarily information about causal relationships. For example,

In a grocery store, the correlation between customer purchases of bread and the purchases of milk may be high and may be a frequently occurring data pair (association).

This does not necessarily indicate a causal relationship; i.e.,

Someone's purchase of bread is caused by their purchase of milk (or the other way around).

Sometimes, high medical attribute correlations have led to questionable causal statements. For example,

A Prussian army physician, Hans Reiter, observed a syndrome (called Reiter's syndrome) that often occurred where the patient also had syphilis. He concluded that Reiter's syndrome was caused by syphilis.

This causal connection description has found its way into many medical textbooks. Unfortunately for this causal connection statement, the incidence of syphilis was high in most armies of that time and almost any disorder, including the common cold, could have been similarly causally connected (by a high correlation) to syphilis. (In the Prussian army, the incidence of Reiter's syndrome was lower than the incidence of either syphilis or common colds.)

Similarly, some years ago,

Epidemiologists found that women taking hormone therapy replacement therapy (HRT) had fewer heart attacks [6]. However follow on controlled studies showed that HRT caused more heart attacks. It turned out that women taking HRT in the epidemiologic study were from higher socioeconomic classes, so they ate and exercised better. Economic class caused both HRT and fewer heart attacks.

Networks (graphs) of causes and effects can be constructed; i.e.,

$$(\alpha \rightarrow \beta, \gamma), (\beta \rightarrow \delta), (\gamma \rightarrow \phi), (\beta \cap \delta \rightarrow \lambda)$$

The question is: How to best represent a causal network if the following characteristics are to be part of it: direction, strength, imprecision, network (graph) structure?

When trying to *precisely* reason about causality, complete knowledge of all of the relevant events and circumstances is needed. As complete knowledge may not be possible, imprecise causal reasoning is needed.

A. An Association Does Not, By Itself, Mean Causality

Associations or correlations do not *necessarily* describe causality. Sometimes, the association might be causal; for example, if

Someone eats salty peanuts, then drinks beer.

or

Someone drinks beer, then becomes inebriated.

there may be a degree of a causal relationship. On the other hand, if

A rooster grows, then the sun rises.

or

Someone wears a 'lucky' shirt, then wins a lottery.

there may not be a causal relationship. Recognizing true causal relationships greatly enhances the decision value of data analysis.

The naïve use of simple associations may lead to errors. Errors might occur; either if causality is recognized where there is no causality; or if the direction of the causal relationship is wrong [7] [8]. For example, if

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

• Is it the following rule?

Our customers are sick, so they buy from us.

• Is it the following complementary rule?

If people use our products, they are likely to become sick.

When making a decision, it is not enough to know that

People both buy our products and are sick.

what is needed is knowledge of what causes what; and, to what degree is something a causal effect of another thing.

B. Granularity

Causality is often granular. This is true both with commonsense reasoning as well as for more formal mathematical and scientific approaches. At a very fine-grained level, the physical world itself may be granular [9]. Commonsense perceptions of causality are often large-grained while the underlying causal structures may be fine-grained.

Larger-grained causal objects are often more imprecise than some of the components that are collected into the larger-grained object. Some components of a larger-grained causal object may be precisely known, while others may be somewhat imprecise, and others unknown. The larger the grain, the greater is the likelihood that there might be missing or unknown components.

How to best evaluate the impreciseness of a larger-grained causal object when the impreciseness of the underlying cascade of components is not clear; perhaps, some form of Type-II fuzzy [10] manipulation might be helpful.

C. Reasoning

Commonsense understanding of the world tells us that we have to deal with imprecision, uncertainty and imperfect knowledge. This is also the case with scientific knowledge of

the world. A difficulty is striking a good balance between precise formalism and commonsense imprecise reality.

Causal relationships exist in the commonsense world; 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 break occurs; or, someone catches the glass before it hits the floor.

Counterfactually, usually (but not always),

Not falling to the floor prevents breakage.

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 α then β* (or, $\alpha \rightarrow \beta$). For example:

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*.

However, 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 swerved and hit it while the car was stopped.

Similarly, negating a causal description often does not work. This is often because an effect can be overdetermined. Also, negative statements are weaker than positive statements as the negative statements can become *overextended*. For example, 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 at all occurs. (There may be no other cars; the other cars brake in time; etc.)

Negative causal relationships are less sure as negative statements are easily overextended; but negative causal relationships are often stated; for example, it is often said that:

Not walking under a ladder prevents bad luck.

Or, usually (but not always), as before

Stopping for a red light avoids an accident.

Perhaps, complete knowledge of all possible factors might lead to a crisp understanding of whether an effect will occur. However, it is unlikely that all possible factors can be known for most situations, as the knowledge of at least some causal effects is imprecise for both positive and negative descriptions. Consequently, causal reasoning models must accommodate inherent ambiguity [13].

When trying to precisely reason about causality, complete knowledge of all of the relevant events and circumstances is needed. In commonsense, every day reasoning, approaches are used that do not require complete knowledge. Often, a *satisficing* [14] paradigm is acceptable.

II. GRANULARITY AND COMPLEXES

When events happen, there are usually other related events. The entire collection of events can be called a complex. A “mechanism” [15] or a “causal complex” [16] [17] is a collection of events whose occurrence or non-occurrence results in a consequent event happening.

Each complex, taken as a whole, often can be reduced to a granule if the grain size is increased. Alternatively, larger complexes often 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 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. One of the events might be used to represent the collection of equal grain sized events; or, a higher-level granule might be specified with the understanding that it will invoke a set of finer-grained events.

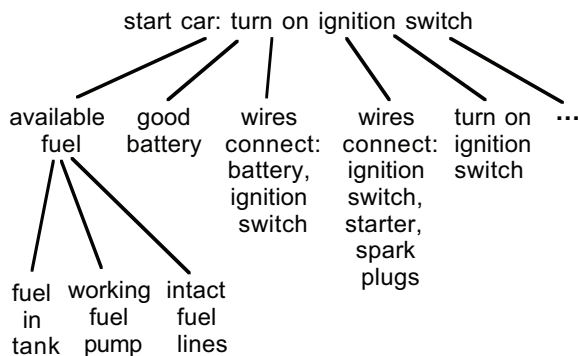


Figure 1. Nested causal complex.

In nested complex granules, the largest-grained event is the sole causal element: *Turn on the ignition switch*. The complex’s other elements represent finer-grains. These elements in turn could be broken down into still finer-grains; for example: “available fuel” can be broken down into:

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

Sometimes, it is enough to know what happens at a large-grained level; at other times it is necessary to know the fined 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.

However,

The automobile mechanic needs to know a finer-grained view of an automobile’s causal complex than does Robin.

Instead of being concerned with all of the fined grained detail, a better approach may be to incorporate granulation using rough sets [18] and/or fuzzy sets to soften the need for preciseness.

In causal reasoning, commonsense reasoning may recognize that a complex collection of elements can be involved causally in a particular effect, even if the precise elements of the complex are unknown. It may not be known what events are in the complex; or, what constraints and laws the complex is subject to. Sometimes, the details underlying an event are known, sometimes not. For example:

A traffic jam might be observed in mid-town Manhattan. A reasonable supposition might be made that the jam was caused by a complex collection of elements. It might be possible to conjecture what were some of the causes; but, it is unlikely to be able to precisely know what were the complete set of actual causes.

Causal complexes may be described using nested granularity. A complex may have several larger-grained elements. In turn, each of the larger-grained elements may be made up of a complex of more fine-grained elements. Recursively, 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 to many fine-grained elements.

When using large-grained commonsense reasoning, people do not always need to know the extent of the underling complexity.

Somebody waiting for a dinner companion in mid-town Manhattan who observed a traffic jam might presume that they might be late because their taxi would most likely be delayed. Only the imprecise large-grained knowledge of the traffic jam would be needed; knowing the more precise finer-grained details of the jam are not be needed.

Not needing to know fine grained detail is also true for situations not involving commonsense reasoning; i.e.:

When designing an electric circuit, designers are rarely concerned with the precise properties of the materials used; instead, they are concerned with the devices functional capabilities and take the device as a larger-grained object.

Hobbs [16] uses first order logic to describe his causal complexes. Pearl [19] develops probabilistic causal networks of directed graphs (DAGs). The causal complexes explicitly considered separately by Hobbs and Pearl have a required structure that may be overly restrictive for commonsense causal understanding:

- 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 may lead to greater savings. However,

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

If all of the events happen, the effect may happen. If some events happen, the effect may still happen. In the commonsense world, 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

An example of restructuring a causal graph by increasing the grain size can be seen in Figures 2a,b. In Figure 2a, the idea is that solid, middle-sized, spherical things all roll are replaced with the concept class representation “ball” of Figure 2b. However, the price of collapsing the variables is a decrease in precision.

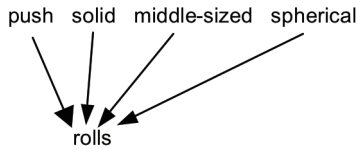


Figure 2a. Variables (solid, middle-sized, spherical) with common effect. [20]

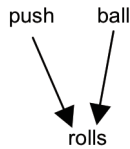


Figure 2b. Variables with a common effect (solid, middle-sized, spherical) collapsed into the concept “ball.” [20]

An issue is how to distinguish between what is in a complex and what is not. Another issue is how to distinguish between the things that deserve to be called “causes” and those that do not. Hobbs [16] suggests that a consideration of causal complexes can be divided into:

- Distinguishing what events are in a causal complex from those outside of it. [21] [22] [15] [11] [19]
- Within a causal complex, recognizing the events that should be identified as causes from those that are not. [23] [24]

A major question concerning complexes is: To what extent can we increase the causal grain size and still have useful causal information? Conversely, can we start with a large-grained causal event and then derive the finer-grained structure? Can we measure and/or control the imprecision involved in changing grain size? If we start with a large-grained structure and resolve it, will our computational complexity burdens be reduced?

III. DIRECTED CAUSAL GRAPH ELEMENTS

The idea of “positive” causation ($\alpha \rightarrow \beta$) is at the commonsense causal reasoning core. Often, a positive causal relationship is represented as a network of nodes and branches with causal direction shown by an arrow.

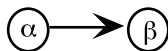


Figure 3. Diagram indicating that α is causally dependent on β .

A. Directed Acyclic Graphs (DAGs)

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 [19]. Silverstein [25] [26] followed a similar approach. From the commonsense causal reasoning view, the various directed graph methods have similar liabilities, specifically:

A.1 Liability: Discrete or continuous data must be reduced to Boolean values. This is an early technique that was and is used in data mining when analyzing market basket data. However, it is essentially flawed. Quantities do matter; some data co-occurrences are conditioned on there being a sufficiency of a co-occurring attribute. Also, some relationships may be non-linear based on quantity [8].

A.2. Liability: There can be no missing data. This is at variance with day-to-day experience. There is often missing data of some sort. Data collection is rarely fully representative and complete. Incremental data is often acquired that is at variance with previously acquired data. A methodology that is not brittle in the face of incompleteness is needed.

A.3. Liability: Causal relationships are not cyclic, either directly or indirectly (through another attribute). This is at variance with our commonsense understanding of the world. Within cyclic dependencies, there are several variants.

A.3.1 Cycles with time lag: There are many commonsense examples where cycles with a time lag are needed.

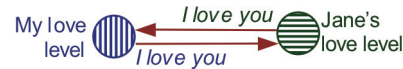


Figure 4. Positive feedback cycle: I tell Jane that I love her. Then, she tells me that she loves me. Then, I tell Jane that I love her more than before. Then, she ... and so forth and so forth. Clearly, the cyclic reinforcement would be substantial.

Some cycles can be reasonably collapsed, some cannot be.

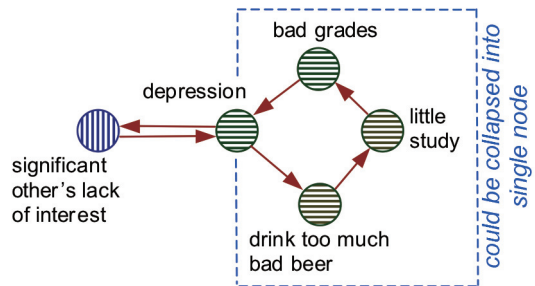


Figure 5. Cyclic relationship that can be collapsed while retaining meaning.

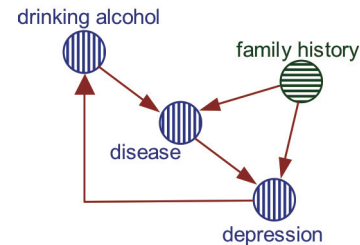


Figure 6. Cyclic relationship that cannot be easily collapsed while retaining meaning.

A.3.2 Concurrent cycles. Another form of a cycle is *joint mutual dependency*. It is possible that there might be mutual dependencies [13]; i.e., $\alpha \rightarrow \beta$ as well as $\beta \rightarrow \alpha$. It seems to be possible that they do so with different strengths. They can be described as shown in the following figure where $S_{i,j}$ represents the strength of the causal relationship from i to j .

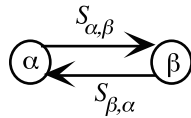


Figure 7. Cyclic relationship: Mutual unequal dependency. [13]

A.3.2.1 Different causal strengths for the same activity, occurring at the same time. 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 understanding and representation. In contrast, Simon [15] and Shoham [27] identify cases where causality is simultaneous.

Mutual causal influences are also our commonsense experience. For example, in the preceding figure, α could be *short men* and β could be *tall women*. If $S_{\alpha,\beta}$ meant the strength of desire for a social meeting, it might be that $S_{\alpha,\beta} > S_{\beta,\alpha}$.

A.3.2.2 Different causal strengths for symmetric activities, occurring at different 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.

A.4 Liability: Markov Stationary Condition holds: Probabilities are time independent (DAG requirement). This does not correspond to our commonsense understanding of the world. When one event is dependent on two causal events, if one causal event happens much earlier (or later) than the other event, there may well be a different result. For example, if the goal is to have a successful dinner party, the catered food should be delivered before the guests arrive.

A.5 Liability: The Markov Memoryless Condition holds (DAG requirement). The *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. This makes sense in the example shown in the following figure.

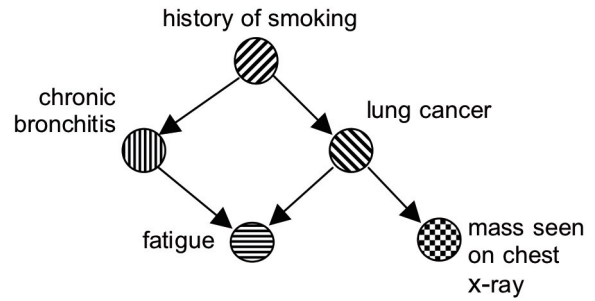


Figure 8. Memoryless Markov condition holds

However, not all commonsense perceptions of causality work this way. Often, we believe that history matters as in the example shown in the following figure.

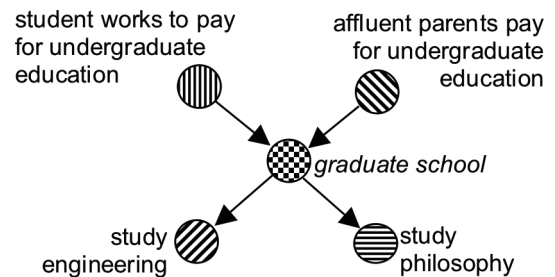


Figure 9. Causality where memory play a part. Memoryless Markov condition does not hold.

B. Markov Models

One possible way of describing the causal relationships is with first order Markov Models. They treat a system as a series of states with specific, constant rate transitions between them. At all times, the system is in exactly one state. (Transitions are considered to be instantaneous.) The only information available is the current state, the allowed transitions, and the probability of these transitions. This means that the system is totally characterized by its current state. None of the past states or transitions have any effect on the transitions out of the current state.

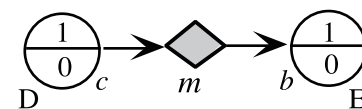


Figure 10. Random Markov model: $c = P(D)$, m = the possibility/probability that when D is present, the causal mechanism brings about E , b = the probability that some other (unspecified) causal mechanism brings about E .

Describing the relationships between the nodes can be complex. One possibility is an extension of the random Markov model, shown in the following figure.

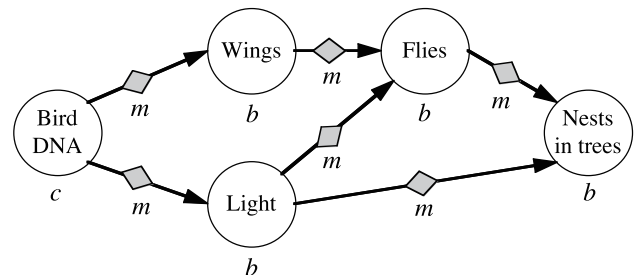


Figure 11. Markov model applied to a "bird" example [28]

Perhaps, fuzzy sets and fuzzy Markov algorithms in particular might be a good way of approaching causality represented by Markov models. Cheok [29] shows how the probabilities in the classic Markov model may be extended to fuzzy possibilities. The generalized hidden Markov Model can be defined with the same type of model parameters, but with a different mathematical basis than the classic one. Fuzzy logic replaces probability theory and this leads to a new definition of the model variables. The structure in terms of states and observations remains the same.

There has been a fair amount of work in fuzzy Markov models [30] [31] [32]. A fair amount of the work has a control theory or medical diagnostic orientation. Whether it can be extended to representing causality is to be seen. While Markov models are intuitively appealing, they have some of the cyclic representational problems as do DAGs. Consequently, they may not be a good solution.

IV. FUZZY COGNITIVE MAPS

A potentially useful form for representing causal structures that overcomes some of the deficiencies of directed acyclic graphs or similar representations is Fuzzy Cognitive Maps (FCMs). They bring their own liabilities in that how to autonomously form FCMs through data mining is unresolved.

Fuzzy Cognitive Maps are fuzzy structures that resemble neural networks. They have powerful and far-reaching consequences as a mathematical tool for modeling complex systems. At first, Axelord [33] used cognitive maps as a formal way of representing social scientific knowledge and modeling decision making in social and political systems [34]. Then, Kosko enhanced cognitive maps considering fuzzy values for them [35] [36] [37] [38].

Indeterminacy and unpredictability of life affect causal descriptions almost as much as the determined factors. It is a major handicap in mathematical modeling if we are only able to give precise values for known concepts and relationships between concepts, thereby presenting onto ourselves a skewed view. Fuzzy theory can offer more choices. Mendel [10] and Kandasamy [39] have introducing indeterminacy into Fuzzy theory; Mendel uses Type-2 Fuzzy Logic while Kandasamy uses Neutrosophic structures.

A FCM describes the behavior of a system in terms of concepts; each concept represents a state or a characteristic of the system. A FCM illustrates the system by a graph showing cause and effect between concepts. FCMs model the world as a collection of classes and causal relations between classes. FCMs are fuzzy signed directed graphs with feedback. The directed edge W_{ij} from causal concept C_i to concept C_j measures how much C_i causes C_j . The time varying concept function $C_i(t)$ measures the non-negative occurrence of some fuzzy event.

A Fuzzy Cognitive Map consists of nodes that represent concepts and arcs (edges) between concepts [40]. Each concept represents a characteristic of the system; in general it stands for events, actions, goals, values, trends of the system. A number A_i that represents its value characterizes each concept and it results from the transformation of the value of the

system's variable, for which this concept stands, in the interval $[-1,1]$. The graph's edges are the casual influences between the concepts.

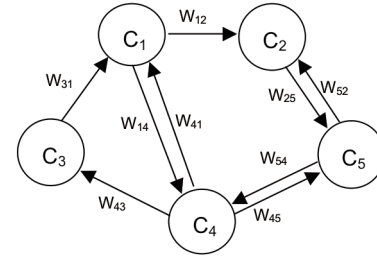


Figure 12. Example of a simple Fuzzy Cognitive Map

The edges W_{ij} takes values in the fuzzy causal interval $[-1, 1]$. The weights of the arcs between concept C_i and concept C_j could be positive ($W_{ij} > 0$) which means that an increase in the value of concept C_i leads to the increase of the value of concept C_j , and a decrease in the value of concept C_i leads to the decrease of the value of concept C_j . Or there is negative causality ($W_{ij} < 0$) which means that an increase in the value of concept C_i leads to the decrease of the value of concept C_j and vice versa. $W_{ij} = 0$ indicates no causality.

State vectors C are repeatedly passed through the FCM connection matrix W , thresholding or non-linearly transforming the result after each pass. The limit cycle or fixed-point inference summarizes the joint effects of all the interacting fuzzy knowledge.

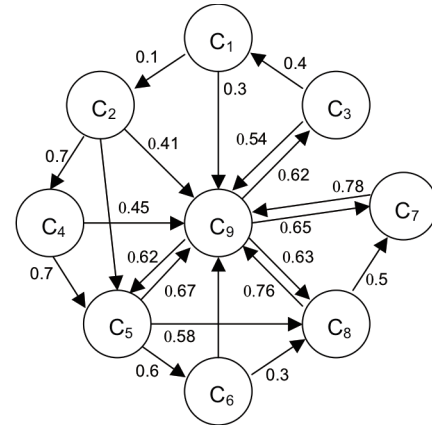


Figure 13. A urinary bladder tumor grading model [41]

The previous figure shows how a medical FCM application; where $C_n, n=1...8$ represent histological features using five positive linguistic variables depending on the characteristics of each particular concept. C_9 represents the target, tumor grade. The concepts are:

- C_1 : Cell distribution: Even, clustered
- C_2 : Cell size: Uniform, pleomorphic
- C_3 : Cell number: Numerous, variable
- C_4 : Cytoplasm: Homogeneous, variable
- C_5 : Nuclei: Uniform, irregular, very irregular, bizarre
- C_6 : Nucleoli: Inconspicuous, evident, prominent
- C_7 : Necrosis: Inconspicuous, frequent
- C_8 : Mitoses: Absent-rate, occasional, numerous
- C_9 : Degree of tumor grade

The value of a node reflects the degree to which the concept is active in the system at a particular time. This value is a function of the sum of all incoming edges multiplied and the value of the originating concept at the immediately preceding state. The threshold function applied to the weighted sums can be fuzzy in nature. Concept values are expressed on a normalized range denoting a degree of activation rather than an exact quantitative value.

The mathematical model consists of $1 \times n$ state vector A that includes the values of the n concepts and a $n \times n$ weight matrix W that gathers the weights W_{ij} of the interconnections between the n concepts of the FCM. The matrix W has n rows and n columns where n equals the total number of distinct concepts of the FCM and the matrix diagonal is zero since it is assumed that no concept causes itself. The value of each concept is influenced by the values of the connected weighted concepts and by its previous value. The value A_i for each concept C_i is calculated by:

$$A_i f \left(\sum_{\substack{j=1 \\ j \neq i}}^n A_j W_{ji} \right) + A_i^{old}$$

where A_i is the activation level of concept C_i at time $t+1$, A_j is the activation level of concept C_j at time $t+1$, A_j is the activation level of concept C_i at time t , and W_{ji} is the weight of the interconnection between C_j and C_i , and f is a threshold function.

$$A_{new} = f(A_{old} \circ W) + A_{old}$$

The new state vector A_{new} is computed by multiplying the previous state vector A_{old} by the weight matrix W . The new vector shows the effect of the change in the value of one concept in the entire Fuzzy Cognitive Map.

The simplest FCMs act as asymmetrical networks of threshold or continuous concepts and converge to an equilibrium point or limit cycles. FCMs have nonlinear structure of their concepts and differ in their global feedback dynamics [42] [38].

There are several unsupervised learning algorithms that might be used to reach convergence, or at least convergence within an acceptable tolerance [43] [44] [45] [46] [47] [48] [49]; the algorithms can be comparatively analyzed [50]; however, neither particular methods or a comparative analysis of the methods are within the scope of this paper. The FCM system model takes the initial values of the concepts from the real system and is free to interact. The interaction happens because of change in the value of one or more concepts [51]. The interaction continues until the model:

- Reaches equilibrium at a fixed point, with the output concept values stabilizing.
- Exhibits limit cycle behavior, with the concept values falling in a loop.
- Exhibits chaotic behavior in a non-deterministic way.

Causal feedback loops can occur in FCMs. FCM feedback allows causal adaptation. FCM feedback forces the abandonment of classic graph search, forward and backward

chaining. Instead, the FCM is a dynamical system and take its equilibrium behavior as a forward-evolved inference. A FCM has the ability to specify a complex model, shows linear or non-linear relations and allows causal propagation [52].

The development of a FCM often occurs within a group context. Experts can pool or fuse their knowledge by adding the underlying FCM causal matrices [43]. The obvious advantage is that a causal relationship can be described in stages. If some form of data mining is discovering the causal relationships, it could be a highly desirable incremental method. Another claimed advantage [34] is that combining incomplete, conflicting opinions of different experts may cancel out the effect of oversight, ignorance and prejudice; although this would seem to require a fairly large number of overlapping relationships so a useful average relationship would be developed.

V. EPILOGUE

The target of many studies in the health sciences is the discovery of cause-effect relationships among observed variables of interest, for example: treatments, exposures, preconditions, and outcomes. Causal modeling and causal discovery are central to medical science. In order to algorithmically consider causal relations, the relationships must be placed into a representation that supports manipulation. An algorithmic way of handling causal imprecision is needed. There are several needs of a causal model. The needs are:

- Represent causal direction
- Represent degree of causal effect
- Represent causal imprecision
- Accommodate changes in grain size
- Describe complexes
- Support cycle models of all times
- Be time varying
- Not be restricted by Markov conditions

The most widespread causal representation is directed acyclic graphs (DAGs). However, DAGs are severely limited in what they can represent. This paper considered the needs of commonsense causality and suggested Fuzzy Cognitive Maps as an alternative to DAGs. Fuzzy Cognitive Maps appear to be a good approximation of complex casual complexes. They appear to be a comprehensive representation. Their primary difficulty lies with establishing their initial values.

REFERENCES

- [1] G. F. Cooper and C. You, "Causal Discovery From A Mixture Of Experimental And Observational Data," presented at UAI, 116-125, 1999.
- [2] F. Heylighen, "Causality As Distinction Conversation: A Theory Of Predictability, Resersibility, And Time Order," *Cybernetics And Systems*, vol. 20, pp. 361-384, 1988.
- [3] J. Pearl, "Causal Inference In The Health Sciences: A Conceptual Introduction," *Health Services And Outcomes Research Methodology*, vol. 2, pp. 189-220, 2001.
- [4] L. A. Zadeh, "The Concept Of A Linguistic Variable And Its Application To Approximate Reasoning - I," *Information Sciences*, vol. 8, pp. 199-249, 1975.
- [5] L. A. Zadeh, "Fuzzy Sets," *Information and Control*, vol. 8, 1965.
- [6] M. Ridley, "When Scientists Confuse Cause And Effect," in *Wall Street Journal*, 2011, pp. C4.

- [7] S. Brin, R. Motwani, and C. Silverstein, "Beyond Market Baskets: Generalizing Association Rules To Correlations," presented at ACM SIGMOD International Conference, SIGMOD'97, 265-276, SIGMOD Record, 1997.
- [8] L. J. Mazlack, "Causality Recognition For Data Mining In An Inherently Ill Defined World," presented at 2003 BISC FLINT-CIBI International Joint Workshop On Soft Computing For Internet And Bioinformatics, Berkeley, California, 2003.
- [9] L. Smolin, "Atoms of Space And Time," in *Scientific American*, 2004, pp. 66-75.
- [10] J. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. New Jersey: Prentice Hall, 2000.
- [11] C. L. Ortiz, "A Commonsense Language For Reasoning About Causation And Rational Action," *Artificial Intelligence*, vol. 108, pp. 125-178, 1999.
- [12] L. A. Zadeh, "From Computing With Numbers To Computing With Words - From Manipulation Of Measurements To Manipulation Of Perceptions," *IEEE Transactions on Circuits and Systems*, vol. 45, pp. 108-119, 1999.
- [13] L. J. Mazlack, "Commonsense Causal Modeling In The Data Mining Context," presented at IEEE International Conference on Data Mining series (ICDM), Melbourne, Florida, 2003.
- [14] H. A. Simon, "A Behavior Model Of Rational Choice," *Quarterly Journal of Economics*, vol. 59, pp. 99-118, 1955.
- [15] H. A. Simon, "Nonmonotonic Reasoning and Causation: Comment," *Cognitive Science*, vol. 15, pp. 293-300, 1991.
- [16] J. R. Hobbs, "Causality," presented at Common Sense 2001, Fifth Symposium on Logical Formalizations of Commonsense Reasoning, 145-155, New York University, New York, 2001.
- [17] J. R. Hobbs, "Causality And Modality: The Case Of 'Would'," *Journal of Semantics*, 2003.
- [18] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht, Netherlands: Kluwer Academic Publishing, 1991.
- [19] J. Pearl, *Causality*. New York, NY: Cambridge University Press, 2000.
- [20] C. Glymour, *The Mind's Arrows, Bayes Nets and Graphical Causal Models in Psychology*. Cambridge, Massachusetts: MIT Press, 2001.
- [21] D. Lewis, "Causation," in *Philosophical Papers II*: Oxford University Press, 1986, 159-213.
- [22] H. A. Simon, "On The Definition Of The Causal Relation," *Journal of Philosophy*, vol. 49, pp. 517-528, 1952.
- [23] J. Mackie, "Causes and Conditions," *American Philosophical Quarterly* 2, pp. 245-64, 1965.
- [24] Y. Shoham, "Nonmonotonic Reasoning And Causation," *Cognitive Science*, vol. 14, pp. 213-252, 1990.
- [25] C. Silverstein, S. Brin, and S. Mani, "Beyond Market Baskets: Generalizing Association Rules To Dependence Rules," *Data Mining And Knowledge Discovery*, vol. 2, pp. 39-68, 1998.
- [26] C. Silverstein, S. Brin, R. Motwani, and J. Ullman, "Scalable Techniques For Mining Causal Structures," presented at International Conference on Very Large Databases, 594-605, New York, NY, 1998.
- [27] Y. Shoham, "Remarks on Simon's Comments," *Cognitive Science*, vol. 15, pp. 301-303, 1991.
- [28] B. Rehder, "A Causal-Model Theory of the Conceptual Representation of Categories," presented at FLAIRS 2002 Workshop on Causality, Pensacola, Florida, 2002.
- [29] A. Cheok, S. Chevalier, M. Kaynak, K. Sengupta, and K. Chung, "Use of a Novel Generalized Fuzzy Hidden Markov Model for Speech Recognition," presented at IEEE International Conference on Fuzzy Systems, Melbourne, Australia, 2002.
- [30] L. Jeanpierre and F. Charpillet, "Hidden Markov Models for Medical Diagnosis," presented at International Workshop on Enterprise Networking and Computing in Health Care Industry, HEALTH-COM'2002, Nancy, France, 2002.
- [31] M. Leuschen, I. Walker, and J. Cavallaro, "Robot Reliability Through Fuzzy Markov Models," presented at IEEE Annual Reliability and Maintainability Symposium, 209-214, Anaheim, CA, 1998.
- [32] L. A. Zadeh, "Maximizing Sets And Fuzzy Markov Algorithms," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 28, pp. 9-15, 1998.
- [33] R. Axelord, *Structure Of Decision: The Cognitive Maps Of Political Elites*. Princeton, New Jersey: Princeton University Press, 1976.
- [34] J. Aguilar, "A Survey About Fuzzy Cognitive Maps Papers," *International Journal Of Computational Cognition*, vol. 3, pp. 27-33, 2005.
- [35] B. Kosko, "Fuzzy Cognitive Maps," *Int. J. Man-Machine Studies*, vol. 24, pp. 65-75, 1986.
- [36] B. Kosko, *Neural Nets And Fuzzy Systems, A Dynamic System Approach To Machine Intelligence*. New Jersey: Prentice Hall, 1992.
- [37] B. Kosko and J. Dickerson, "Fuzzy Virtual Worlds," *AI Expert*, pp. 25-31, 1994.
- [38] B. Kosko, *Fuzzy Engineering*. New Jersey: Prentice Hall, 1997.
- [39] W. B. V. Kandasamy and F. Smardandache, *Fuzzy Cognitive Maps And Neutrosophic Cognitive Maps*. Phoenix: Xiquan, 2003.
- [40] C. D. Stylios, V. Georgopoulos, and P. P. Groumpos, "The Use Of Fuzzy Cognitive Maps In Modeling Systems," presented at The 5th IEEE Mediterranean Conference on Control and Systems (MED5), paper 67 (CD-ROM), Paphos, Cyprus, 1997.
- [41] E. I. Papageorgiou, P. Spyridonos, C. Stylios, P. Ravazoula, P. P. Groumpos, and G. Nikiforidis, "A Fuzzy Cognitive Map Model For Grading Urinary Bladder Tumors," presented at 5th Int. Conf. in Neural Networks & Experts Systems in Medicine & Healthcare 1st Int. Conf. in Computational Intelligence in Medicine & Healthcare, NNESMED/CIMED Conference, 6, 2003.
- [42] B. Kosko, "Fuzzy Associative Memory Systems," in *Fuzzy Expert Systems*, A. Kandel, Ed. Boca Raton, Florida, 1992, 135-162.
- [43] R. Taber, R. R. Yager, and C. M. Helgason, "Quantization Effects on the Equilibrium Behavior of Combined Fuzzy Cognitive Maps," *International Journal of Intelligent Systems*, vol. 22, pp. 181-202, 2007.
- [44] E. I. Papageorgiou, C. Stylios, and P. P. Groumpos, "An Activation Hebbian Learning Algorithm To Train Fuzzy Cognitive Maps," *International Journal of Approximate Reasoning*, vol. 37, pp. 219-249, 2004.
- [45] Y. Cai, C. Miao, A.-H. Tan, and Z. Shen, "Context Modeling With Evolutionary Fuzzy Cognitive Map In Interactive Storytelling," presented at Fuzz-IEEE, 2320-2325, Hong Kong, China, 2008.
- [46] A. Konar and U. K. Chakraborty, "Reasoning And Unsupervised Learning In A Fuzzy Cognitive Map," *Information Sciences*, vol. 170, pp. 419-441, 2005.
- [47] S.-J. Li and R.-M. Shen, "Fuzzy Cognitive Map Learning Based On Improved Nonlinear Hebbian Rule," presented at Proceedings Of International Conference On Machine Learning And Cybernetics 4, 2301-2306, 2004.
- [48] E. I. Papageorgiou and P. P. Groumpos, "A Weight Adaptation Method For Fuzzy Cognitive Maps To A Process Control Problem," in *Lecture Notes In Computer Science*, vol. 3037. Berlin: Springer, 2004.
- [49] W. Stach, L. Kurgan, W. Pedrycz, and M. Reformat, "Genetic Learning Of Fuzzy Cognitive Maps," *Fuzzy Sets And Systems*, vol. 143, pp. 371-401, 2005.
- [50] A. K. Tsadiras, "Comparing The Inference Capabilities Of Binary, Trivalent And Sigmoid Fuzzy Cognitive Maps," *Information Sciences*, vol. 178, pp. 3880-3894, 2008.
- [51] E. I. Papageorgiou, C. Stylios, and P. P. Groumpos, "Unsupervised Learning Techniques For Fine-Tuning Fuzzy Cognitive Map Causal Links," *Human-Computer Studies*, vol. 64, pp. 727-743, 2006.
- [52] C. D. Stylios and P. P. Groumpos, "The Challenge Of Modeling Supervisory Systems Using Fuzzy Cognitive Maps," *Journal of Intelligent Manufacturing*, vol. 9, pp. 339-345, 1998.