Intensity Reasoning by Constraint Propagation Based on Causal Relationships

Koichi YAMADA
Planning and Management Science, Nagaoka University of Technology
1603-1, Kami-tomioka, Nagaoka, Niigata, 940-21 JAPAN
Tel: +81-258-47-9359, Fax: +81-257-47-9350, Email: yamada@kjs.nagaokaut.ac.jp

Abstract
Many events in the world occur with some quantity that shows a level of the occurrence. This paper discusses reasoning with the normalized level of the occurrence, which we call intensity, and proposes intensity reasoning by constraint propagation based on causal relationships.

The knowledge used in the reasoning is given by a directed acyclic graph called causal network. Each node in the network expresses an event with an intensity in [0,1], and each arc between two nodes does a causal relation defined by a function that gives the relation between intensities of those nodes. The reasoning is conducted by constraint propagation to derive a range of intensity of an arbitrarily chosen node when those of some other nodes are given.

1. Introduction
Knowledge about relationships among objects or events plays an important role to comprehend the structures of the world, and to derive unknown information from the known. Especially, causality is one of the most basic and important relations, because it is “a ubiquitous notion in man’s conception of his environment”[1], “a language with which one can talk efficiently about certain structures of relevance relationships, with the objective of separating the relevant from the superfluous”[1], and “an essential element of human understanding of all phenomena in the world”[2].

In spite of the important notion of the causality, research done so far for reasoning based on causalities is not sufficient. Of course, there are a few theories that deal with reasoning based on them. Probabilistic network[1] and qualitative reasoning[2] are well-known reasoning theories based on causalities. Some of abduction theories are also based on causal relationships[3]. However, they do not cover all types of reasoning with causalities necessary for the real world problems.

One issue we have to consider is the direction of reasoning. One popular way of reasoning with causalities is deduction which employs them in the forward manner. It derives some effects from causes. The other is abductive reasoning which uses causalities in the backward manner, and derives some causes as explanations of given effects. However, it seems that most of real world problems which needs intelligence can NOT be solved only by deductive reasoning nor by abductive reasoning. When human beings face such difficult problems, they solve them by performing the two types of reasoning simultaneously or alternately. This way of reasoning, however, is possible only in probabilistic network among those mentioned above.†

Another issue is the case we have to consider “intensities”[4] of events. Intensity is a normalized quantity pertaining to an event when it occurs. “Degree of nastiness” for event “nasty smell”, “normalized rainfall” for event “rain”, and “normalized index of recession” for event “recession” are typical examples of them. These intensities are considered to express a kind of fuzziness. However, different from probability and possibility, intensity has hardly been dealt with in a framework of reasoning.††

This paper proposes a way of reasoning thorough propagation of intensity constraints in a directed acyclic graph called causal network. Each node in the network expresses an event with an intensity in [0,1], and each arc between two nodes does a causal relation defined by a function that gives the relation between intensities of those nodes. The reasoning, which includes deduction and abduction, is conducted to derive an intensity constraint of an arbitrarily chosen node when those of some other nodes are given.

2. Consistent, Supportive, Contradictive
Let X, Y be two events that have a causal relationship,

† Though hypothetical reasoning seems to include both deductive and abductive reasoning, the goal that should be proved and possible hypotheses are given a priori.
†† Quantity has been mainly dealt with by mathematical models. Though qualitative reasoning and fuzzy reasoning can be used when mathematical models are unavailable, there are some problems for reasoning with causalities. The main goal of fuzzy reasoning is approximation of nonlinear function by piece-wise-interpolation, while qualitative reasoning has a difficulty to avoid combinational explosion of possible situations.
and suppose that cause $X_i$ is arising with intensity $x_i \in [0, 1]$. In this case, we assume that intensity $y_j^*$ of effect $Y_j$ caused by $X_i$ is given by the next equation:

$$y_j^* = c_{ij}(x_i),$$

where $c_{ij}$ is a non-decreasing continuous function that defines the causality, and satisfies $c_{ij}(0)=0$ and $0 \leq c_{ij}(x_i) \leq 1$.

When $Y_j$ can be caused by other events as well as $X_i$, and those event may be arising, intensity $y_j$ of $Y_j$ has a value between $c_{ij}(x_i)$ and 1.0, if there is no canceling effect among causes.

$$y_j \in [y_j^*, 1] = [c_{ij}(x_i), 1].$$

(2)

On the contrary, when $y_j$ is given, intensity $x_i^*$ of $X_i$ that derives $y_j$ or the possible largest intensity is given by the next equation:

$$x_i^* = d_{ij}(y_j) = \begin{cases} 0 & \text{if } y_j \in [0, c_{ij}(1)] \\ x_i & \text{otherwise} \end{cases}$$

(3)

When $Y_j$ may be caused by another event, intensity $x_i$ should satisfy the following:

$$x_i \in [0, x_i^*] = [0, d_{ij}(y_j)].$$

(4)

When $x_i$, $y_j$, and $c_{ij}(*$) are all given, we say that “$x_i$ and $y_j$ are consistent” if $y_j \geq c_{ij}(x_i)$, and “contradictive” if $y_j < c_{ij}(x_i)$.

Furthermore, $x_i$ supports $y_j$ if $y_j = c_{ij}(x_i)$.

3. Propagation of Intensity Constraints

3.1 Forward Propagation

When $X_i$ and $Y_j$ have a causality $c_{ij}$, and a constraint of $x_i$ is given by $x_i \in [x_i^S, x_i^L]$, it is possible to propagate the constraint to $Y_j$. The propagation, called FP (Forward propagation), has two ways - WFP (weak FP) based on consistency and SFP (strong FP) based on support. WFP and SFP are defined as follows, respectively.

$$y_j^{CN} = \{ y_j \mid y_j \geq c_{ij}(x_i), x_i \in x_i \} = \{ c_{ij}(x_i^S), 1 \}. $$

(5)

$$y_j^{SP} = \{ y_j \mid y_j = c_{ij}(x_i), x_i \in x_i \} = \{ c_{ij}(x_i^S), c_{ij}(x_i^L) \}. $$

(6)

$y_j^{CN}$ and $y_j^{SP}$ are called “consistent constraint of $y_j$ by $x_i$” and “supported constraint of $y_j$ by $x_i$”, respectively.

3.2 Backward Propagation

BP (Backward propagation) is one that propagates constraint $y_j \in [y_j^S, y_j^L]$ of $y_j$ to $X_i$. There are also two kinds of BP - WBP (weak BP) and SBP (strong BP), and they are defined by the followings:

$$x_i^{CN} = \{ x_i \mid y_j \geq c_{ij}(x_i), y_j \in y_j \} = \{ d_{ij}(y_j^L), y_j \}. $$

(7)

$$x_i^{SP} = \{ x_i \mid y_j = c_{ij}(x_i), y_j \in y_j \} = \{ d_{ij}(y_j^S), y_j \}. $$

(8)

$x_i^{CN}$ and $x_i^{SP}$ are “consistent constraint of $x_i$ by $y_j$” and “supporting constraint of $x_i$ by $y_j$”, respectively.

4. Intensity Reasoning with Hierarchical Causal Network

Hierarchical causal network consists of two groups of nodes ($X_i$s and $Y_j$s) and arcs with an arrow that connects $X_i$ to $Y_j$. Nodes $X_i$s and $Y_j$s express events and arcs do causal relationships between them. This section discusses reasoning that derives intensity constraints of all nodes, when some of them are given by the user. We call the nodes whose constraints are given Evidence nodes (E-nodes). The others, not given any constraints by the user, are called Anticipatory nodes (A-nodes), but they also have natural constraints of [0,1].

The reasoning is conducted by the following two steps.

step 1) : Reasoning based on Consistency.

step 2) : Reasoning based on Support.

4.1 Reasoning Based on Consistency

A intensity constraint is given by a interval of possible intensity. The purpose of reasoning based on consistency (RBC) is to remove contradictive range of intensity, which is contradictive to the constraints of the other nodes, from the given constraints. Thus, the results of this reasoning guarantees that intensity within the constraint of any node is consistent with those of all other nodes. However, it does not satisfy supporting and supported constraints generally. In this reasoning, not only intensity constraints of A-nodes but also those of E-nodes are revised, because given constraints of E-nodes may contain contradictive intensity.

RBC is conducted by WFP and WBP. First, let the initial constraints be $x_i^L = x_i^S \in [x_i^S, x_i^L]$, $y_j^L = y_j^S \in [y_j^S, y_j^L]$, where

$x_i^L = y_j^S = 0$, $x_i^L = y_j^S = 1$, if $X_i$ and $Y_j$ are A-nodes. Then, these are replaced by $x_i^0 = [x_i^S, x_i^L]$, $y_j^0 = [y_j^S, y_j^L]$ shown below.

(9)

$$x_i^0 = \begin{bmatrix} x_i^S & x_i^L \end{bmatrix} \quad y_j^0 = \begin{bmatrix} y_j^S & y_j^L \end{bmatrix}$$

The reason why the upper limit of $y_j^1$ is modified is that $y_j \in (\lor_i c_{ij}(1), 1]$ can not be supported by any $x_i$, and
does not give any constraints to other nodes.

Then, WFP is conducted and \( y_i^W \) is replaced by intersection of \( y_i^L \) and all of consistent constraints by \( x_i^W \). After that, WBP is conducted and \( x_i^S \) is replaced by intersection of \( x_i^W \) and all of consistent constraints by the replaced \( y_i^W \). The results are given below.

\[
x_i^W = [x_i^{S0}, \bigwedge_{j} d_{ij}(y_j^{L0}) \land x_j^{L0}] . \tag{11}
\]
\[
y_j^W = [\bigvee_{i} c_{ij}(x_i^{S0}) \lor y_j^{S0}, y_j^{L0}] . \tag{12}
\]

However, if the following conditions are not satisfied, there is no solution and some of \( x_i^W \) and \( y_j^W \) become empty sets.

\[
\forall (i, j), y_j^{L0} \geq c_{ij}(x_i^{S0}) \tag{13}
\]

### 4.2 Reasoning Based on Support

In case of reasoning based on support (RBS), interactions among causalities are more complex than RBC. When RBC is done using WFP from plural \( X_i \) to a \( Y_j \), the final consistent constraint of \( Y_j \) was obtained by intersection of all consistent constraints by \( x_i \). However, supported constraint of \( Y_j \) does not need supports by all \( x_i \). It is sufficient that \( y_j \) is supported by one of \( x_i \) and consistent with all other \( x_i \). The widest constraint \( y_j \) that satisfies the above is given by the next equation.

\[
y_j = \left[ \bigvee_i c_{ij}(x_i^{S}), \bigvee_i c_{ij}(x_i^{L}) \right] , \tag{14}
\]

where \( x_i^{S} \), \( x_i^{L} \) are lower and upper limits of \( x_i \). The propagation by the above equation SFP-CN (SFP on causal networks). The reasoning by SFP-CN corresponds to deductive reasoning mentioned in Introduction.

On the contrary, when we try to obtain \( x_i \)s that satisfy eq. (14) for given \( y_j \), all \( x_i \)s do not have to be \( x_i^{SP} \) given by eq. (8). It is enough that one \( x_i \) is equal to \( x_i^{SP} \) and the others are \( x_i^{CN} \) given by eq. (7). This propagation from \( Y_j \) to \( X_i \) is called SBP-CN (SBP on causal networks), which corresponds to abductive reasoning mentioned in Introduction.

In the following, the process of RBS, which is based on SFP-CN and SBP-CN, is described briefly. The purpose of RBS is to obtain intensity constraints of all A-nodes so that they may satisfy all supporting/supported constraints given by the neighbor nodes. Different from RBC, constraints of E-nodes are fixed and only those of A-nodes are revised.

1. Reasoning based on consistency
   - Conduct RBC, and revise the intensity constraints so that all of them are consistent with one another.

2. Simplification
   - Remove nodes and arcs that do not affect the results of this reasoning in the following ways.
     a) Remove every E-node \( Y_j \) that has no connection with A-nodes \( X_i \).
     b) Remove every E-node \( Y_j \), if it is connected with at least an A-node \( X_i \), and there is an \( X_i \) (A-node or E-node) that satisfies the supporting constraint by \( Y_j \).
     c) Remove every arc that connects E-nodes \( X_i \) and \( Y_j \), if \( Y_j \) is connected with at least an A-node \( X_i \), and there is no \( X_i \) (A-node or E-node) that satisfies the supporting constraint by \( Y_j \).
     d) Remove every arc that connects any nodes \( X_i \) and \( Y_j \), if the next conditions are satisfied.

\[
y_j^{S} \geq c_{ij}(x_i^{T}) \tag{15}
\]

Nodes which are isolated by the above simplification are also removed. The final intensity constraints of removed nodes are those when the nodes are removed. After this simplification, there is no connection between E-nodes \( X_i \) and \( Y_j \).

3. Determinative propagation
   - Determinative propagation is one that can give the final constraint to a node. It is possible when the connection has a pattern shown in Fig. 1 (a) and (b). In these cases, the final constraints \( y_j^{NEW} \) of A-node \( Y_j \), and \( x_i^{NEW} \) of A-node \( X_i \) are determined in the following equations, respectively.

\[
y_j^{NEW} = \left[ \bigvee_i c_{ij}(x_i^{S}), \bigvee_i c_{ij}(x_i^{L}) \right] \cap y_j . \tag{16}
\]
\[
x_i^{NEW} = \bigcap_j \left[ d_{ij}(y_j^{S}), d_{ij}(y_j^{L}) \right] \cap x_i . \tag{17}
\]

In both cases, nodes given the new constraints become E-nodes after the propagation. Then, E-nodes \( Y_j \) are removed, because they cannot propagate any constraints to other nodes.

4. Combinational non-determinative propagation
   - When there is a pattern of Fig. 1 (c) in the remaining network, it is guaranteed that all \( x_i \)s satisfy the consistent constraints \( x_i^{CN} \) by \( y_j \) due to the processing of (1), but \( x_i \) does not satisfy the supporting constraint \( x_i^{SP} \) by \( y_j \) due to the processing of Simplification c). Thus, we choose a \( X_i \) and conduct SBP from \( Y_j \) to the chosen \( X_i \). The constraint of the \( X_i \) is revised as follows:

\[
x_i^{NEW} = \left[ d_{ij}(y_j^{S}), d_{ij}(y_j^{L}) \right] \cap x_i . \tag{18}
\]

After the propagation, E-node \( Y_j \) is removed. \( x_i \)s are still A-nodes, because another constraint may be propagated from another \( Y_j \). Since this propagation is a
shows a E-node. Arcs with X means that such arcs must not exist.

Fig. 1 Connection Patterns in Hierarchical Causal Networks

5. Intensity Reasoning with Directed Acyclic Graph

We are ready to describe the final goal of the paper, which is intensity reasoning with a directed acyclic graph. The reasoning derives an intensity constraint of an arbitrarily chosen A-node called focal, given constraints of several E-nodes. The reasoning is conducted in the following steps:

1) Decompose each E-node and the focal node into several so that they may be terminal nodes. The network may be divided into several subnetworks by the decomposition.

2) Transform each subnetwork into a hierarchical network where all of non-terminal nodes are removed.

3) Apply the reasoning proposed in the previous section to the hierarchical causal network.

The details of 1) and 2) are discussed in the followings.

5.1 Decomposition of E-nodes

In network models where information is propagated from a node to its neighbors, the propagation is blocked by instantiated nodes such as E-nodes, because all of information in them are fixed. Therefore, E-nodes can be decomposed into several other nodes with the same information as that of the original one as shown in Fig. 2. What must be noticed about the decomposition is that the children of the E-node can be isolated from one another (if necessary), while the parents remain connected due to the complex interaction among causal relations in strong propagation. As mentioned in 4.2, only one of $X_1$ and $X_2$ must satisfy the supporting constraint by $X_3$, and the other is OK if it satisfied the consistent constraint by $X_3$. Therefore, we cannot consider $X_1$ and $X_2$ separately.

5.2 Decomposition of the focal node

The focal node is one of A-nodes, and A-nodes cannot be decomposed generally, since the information is not fixed. However, remember that we are going to derive an intensity constraint of just a node, the focal node $F$. If we have no concern with the other A-nodes, the constraint propagated to $F$ from E-nodes do not have to be propagated to its neighbors, as long as the constraint of $F$ does not create any further one on $F$ itself through a feedback loop composed by arcs. We can decompose $F$ to several terminal nodes in the same way as E-nodes, derive constraints of the decomposed nodes, and calculate the intersection of them.

The possibility that $F$ creates a further constraint to itself is considered, only when $F$ is in an undirected...
loop that corresponds to an undirected cycle in a directed graph. Actually, WFP and WBP can be done along the loop, but it does not create any further constraint on F, because the propagated constraints along the loop is always \([0, 1]\). On the other hand, neither SFP or SBP can be conducted along the loop, because they must be applied in the form of SFP-CN and SBP-CN. SFP-CN is always propagated forward, and SBP-CN backward.

5.3 Transformation into Hierarchical Network

When our interest is only in terminal nodes of the network, the network can be transformed into a hierarchical one by removing all of intermediate nodes. The transformed hierarchical network is equivalent to the original one in the sense of the next definition.

[Definition 5.1]

Two causal networks are equivalent when the following three conditions are satisfied:

1) Both have the same number of starting nodes (terminal ones without an arrow) and ending nodes (terminal ones with arrows).
2) When constraints of all starting nodes are given, the derived supported constraints of ending nodes are the same in both of the networks.
3) When constraints of all ending nodes are given, the derived supporting constraints of starting nodes are the same in both of the networks.

In the following, we describe the process to transform a causal network expressed in a directed acyclic graph into the equivalent hierarchical network without proof due to lack of space.

(1) Step 1: Replace serial causalities by one

When there are a series of arcs expressing serial causal relationships, they can be replaced by an arc. Suppose that there are a pair of starting and ending nodes \(X_i\) and \(X_m\), and several intermediate nodes \(X_k\) \((k=2, ..., m-1)\) between them. The arc connecting \(X_k\) and \(X_{k+1}\) has a causal relationship defined by function \(c_k\) \((k=1, ..., m-1)\). Then, the series of causalities can be replaced by one defined by the next equation.

\[
c = c_{m-1} \ldots c_1,
\]

(21)

where \(c\) is a non-decreasing continuous function satisfying \(c(0)=0\) and \(0 \leq c(x) \leq 1\). Applying the above function to eq. (3), we get

\[
d(x_m) = \begin{cases} 
  \bigvee_{x_m \in c(x_1)} x_1, & \text{if } x_m \in [0, c(1)] \\
  1, & \text{otherwise}
\end{cases}
\]

(22)

By replacing all of serial causalities defined by \(c_1, ..., c_m\) by a single causality given in eq. (21), the network is transformed into a hierarchical one equivalent to the original.

(2) Step 2: Replace parallel causalities by one

When the network is multiply connected, that means there is at least an undirected loop mentioned before, the hierarchical network transformed by step 1) has at least a pair of starting and ending nodes which have plural arcs between them as shown in Fig. 3. If constraint \(x_i = [x_i^S, x_i^L]\) is given to \(X_i\) and SFP-CN by eq. (14) is applied, we can get the following supported constrain of \(Y_j\).

\[
y_j = \left[ \bigvee_{h} c_{ij}^h(x_i^h), \bigvee_{h} c_{ji}^h(x_j^h) \right].
\]

(23)

where \(c_{ij}^h\)'s are functions defining arcs connecting \(X_i\) and \(Y_j\) in parallel. The above equation shows us that the parallel arcs can be replaced by one defined by the next function.

\[
c_{ij}(x_i) = \bigvee_{h} c_{ij}^h(x_i)
\]

(24)

In this case, the function corresponding to eq. (3) is given as follows:

\[
d_{ij}(y_j) = \begin{cases} 
  \bigvee_{y_j = \bigvee_{h} c_{ij}^h(x_i)} x_i, & \text{if } y_j \in [0, \bigvee_{h} c_{ij}^h(1)] \\
  1, & \text{otherwise}
\end{cases}
\]

(25)

where \(d_{ij}^h\)'s are functions given by applying \(c_{ij}^h\)'s to eq. (3).

By following step 1, 2) described above, we can transform an arbitrary causal network into the equivalent hierarchical causal network.

6. Illustrative Example

Finally, we show an illustrative example that demonstrates the capability of the reasoning. Fig. 4 gives a part of knowledge expressed in a causal network. Functions defining causal relations are given by \(x_j = c_{ij}(x_i) = r_{ij} x_i\) using the value \(r_{ij}\) attached to each arc.

The network has three E-nodes -- Forestry-brisk

![Figure 3: Nodes connected by plural arcs](image-url)
Bears-appear-in-habitation-area [0.5, 0.7], and Patients-of-hay-fever-decreasing [0.6, 0.8] -- all of which are already terminal nodes. Then, assume we want to know if monkeys in mountain suffered from food shortage or not this winter (the focal node is Monkeys-food-shortage).

First, the focal node is decomposed into two; one connecting to its three parents, the other to its child. Then, the network is transformed into the hierarchical one shown in Fig.5. After that, the reasoning described in Section 4 is applied, and two patterns of solutions shown also in Fig.5 are obtained. The final results are obtained by taking intersection of constraints of Monkeys-food-shortage (1) and (2). In this case, both results happen to be the same, and it says that monkeys suffered from a serious food shortage (its seriousness is between 0.67 to 1) this winter.

What should be noticed here is that the correctness of the obtained constrains is guaranteed only for the focal node. Recall that we could decompose the focal node, because we had no concern with the other A-nodes. Let us examine the result of “Crop-plunder”, for example. The constraints obtained are [0.4, 0.9] in both of solutions, but there are not correct. This is because the strongest causality between “Cool summer” and “Monkeys-food-shortage” is given by the path of “Cool-summer -> Bad-harvest-in-forest -> Monkeys-food-shortage -> Crop-plunder” in Fig.4, while this path is not given in Fig.5 due to the decomposition of Monkeys-food-shortage.

7. Conclusion

The paper proposed Intensity Reasoning which reasons intensities of events by constraint propagation based on causal relationships. The features of this reasoning is 1) it can reason not only occurrence of events but also quantities pertaining to the events, and 2) it includes both deductive and abductive reasoning which must be used simultaneously or alternately to solve complex problems that needs intelligence. Therefore, this reasoning could be applicable to many real world problems that needs analysis based on causalities.

REFERENCES


Koichi Yamada received his MS and Dr. Eng. in Systems Science from Tokyo Institute of Technology. He is an associate professor in Department of Planning & Management Science at Nagoya Univ. of Tech.