Université de Tunis
Institut Supérieur de Gestion de Tunis
Ecole Doctorale Sciences de Gestion

Mémoire de stage

Mémoire en vue de l'obtention du Master, spécialité
« Informatique »

Présentée par

Marwa El abri

Bayesian Network and Cognitive Map for learning and knowledge visualization

Sous la direction de :

M.	Philippe Leray	Professeur à l'Université de Nantes	(Directeur)
M.	Stéphane Loiseau	Professeur à l'Université d'Angers	(Directeur)
Mme.	Nahla Ben Amor	Professeur à l'Université de Tunis	(Co-Directeur)

Mémoire de stage préparée au
Laboratoire d'Informatique de Nantes Atlantique
Ecole Polytechnique de l'université du Nantes
La Chantrerie - rue Christian Pauc
BP 5069 - 44306 Nantes Cedex 3

Remerciements

0.1 Acknowledgement

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

A special gratitude I give to my supervisor **Pr. Philippe Leray**, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project especially in writing this report. I would also like to acknowledge with much appreciation his crucial role, because he gave me the golden opportunity to do this wonderful project, which also helped me in doing a lot of research and I came to know about so many new things. Many thanks to him for introducing me to an interesting research area and inviting me to work in France, where my thesis was totally achieved. My association with him has helped me grow not only as a researcher but also as an individual.

I have to appreciate the guidance given my supervisor **Pr. Stéphane Loiseau** and his research team : **Dr. Davit Genest**, **Dr. Béatrice Duval**, and **Dr. Laurent Garcia**, as well as the panels especially in our project presentation that has improved our presentation skills thanks to their comment and advices.

A thank you to **Dr. Nahla Ben amor** who introduced me to Graphics models, and whose enthusiasm for the "underlying structures" had lasting effect.

I would have not finished this thesis without the support of my family especially my mother who has always been there for me whenever I need them, the encouragement they give to keep me going and their love and prayer to empower me that never fails all the time. Thank you.

Last and not least, I would also like to thank my friends who helped me a lot in finalizing this project within the limited time frame.

Abstract

To solve a given decision problem, there is often need to firstly provide knowledge and judgment analysis of the problem. This task is not easy, especially today in the field of artificial intelligence, the means and the models of visualization and reasoning are numerous and they are becoming more sophisticated. So the difficulty is how to choose the right model to make the right decision. This work is mainly composed of two parts : the first part aims to study similarities and differences between three models (CM,QPN,CBN)and propose firstly and principally a process for incremental construction of CBN starting from an initial modeling CM through QPN, in order to take advantage of each model. Then we propose some semantics extensions. In the second part ; we happened to find links between semantics of QPN and coefficients of log-linear model to fill the probabilities tables of CBN.
Key-words :Cognitive Map, Causal Bayesian Network, Qualitative Probabilistic Network, reasoning, visualization, log-linear model.

Résumé

Pour résoudre un problème de décision donnée, il est souvent nécessaire de prévoir d'abord la connaissance et l'analyse du jugement du problème. Cette tâche n'est pas facile, surtout aujourd'hui dans le domaine d'intelligence artificielle les moyens et modèles de visualisation et de raisonnement sont très nombreux et ils sont de plus en plus évolués. Donc la difficulté c'est comment choisir le bon modèle pour prendre la bonne décision. Ce travail est principalement composé de deux parties : la première partie vise à étudier les similitudes et les différences entre les trois modèles (CM, QPN, CBN) et de proposer d'abord et avant tout un processus de construction progressive de CBN à partir d'une modélisation CM initial à travers un QPN, afin de profiter des avantages de chaque modèle. Ensuite, nous proposons des extensions sémantiques. Dans la deuxième partie, nous avons arrivé à trouver les liens entre la sémantique du QPN et les coefficients du modèle log-linéaire pour remplir les tables de probabilités de CBN.
Mots-clés : Carte cognitive, réseau Bayésien causal, réseau qualitative propabiliste, raisonnement, visualisation, model log-linéaire.

Table des matières

0.1 Acknowledgement iii

Table des matières **vi**

Table des figures **vii**

Liste des tableaux **viii**

1 Graphical knowledge-based systems **5**
- 1.1 Cognitive map 6
  - 1.1.1 Introduction 6
  - 1.1.2 Graphical representation 7
  - 1.1.3 Reasoning with a cognitive map 10
  - 1.1.4 Construction method 10
  - 1.1.5 Extension of cognitive maps 12
- 1.2 Causal Bayesian Networks 13
  - 1.2.1 Introduction 13
  - 1.2.2 Definitions 13
  - 1.2.3 Reasoning with causal Bayesian network 15
  - 1.2.4 Construction of causal Bayes Nets 20
- 1.3 Qualitative Probabilistic Networks 22
  - 1.3.1 Introduction 22
  - 1.3.2 Definition 22
  - 1.3.3 Qualitative influence and synergies 22
  - 1.3.4 The properties of symmetry, transitivity and composition 25
  - 1.3.5 Intercausal influence 25
  - 1.3.6 Reasoning with QPN 26
  - 1.3.7 Extension of QPNs 28
- 1.4 Ontology 29
  - 1.4.1 Introduction 29
  - 1.4.2 Definition 29
  - 1.4.3 Use of ontology 30
  - 1.4.4 construction of an ontology 31
- 1.5 Comparative study 31
- 1.6 Existing bridges 32
  - 1.6.1 Interaction between CM & Causal BNs 32
  - 1.6.2 Interaction between CM & QPNs 33
  - 1.6.3 Interaction between CM & Ontology 33
  - 1.6.4 Interaction between BNs & QPNs 34
  - 1.6.5 Interaction between BNs & Ontology 34

2 Incremental design of Graphical knowledge-based systems 37
2.1 Description of the construction process 37
2.1.1 Phase I : Initial CM 37
2.1.2 Phase II : from CM to QPN 40
2.1.3 Phase III : from QPN to CBN 46
2.1.4 Phase IV : from CBN to QPN 48
2.2 process extension 48
2.2.1 General description of the principle 48
2.2.2 Hierarchy concept : helps in visualization 49
2.2.3 Semantic relations : help in construction 50

3 Estimation of probabilities 51
3.1 Log-linear model approach 51
3.1.1 Links between $\mu - terms$ and CPTs 52
3.1.2 Links between $\mu - terms$ and semantic of QPN 57

Conclusion 59

A Annexes 61
A.1 The d-separation criterion 62
A.2 The representation hypothesis 63

Bibliographie 65

Table des figures

1.1 The evolution of work on cognitive maps. 7
1.2 Example of Cognitive Map. 9
1.3 Example of Fuzzy Cognitive Map. 12
1.4 Example of Bayes Networks. 14
1.5 An altered graph after mutilation 17
1.6 Example of CBN before and after mutilation 18
1.7 The augmented graph G_a. 18
1.8 negative product synergy between A and T with regard to the value b for node B. 25
1.9 Example of ontology. 30

2.1 Schematic overview of the approach. 38
2.2 An example of the procedure for deriving an initial Cognitive Maps, result of phaseI. 40
2.3 Distinguishing between deductive and abductive reasoning 42
2.4 Distinguishing between direct and indirect relations 42
2.5 Simple QPN obtained from CM described in Figure 2.4 . . . 43

2.6 QPN with strength impact,possible improvment of QPN in Figure 2.5 44
2.7 Interval QPN,possible improvment of QPN in Figure 2.5 . . 44
2.8 QPN with product synergies,possible improvment of QPN in Figure 2.5 45
2.9 Example of result of the transformation of qualitative influences to table of probabilities 47
2.10 The Ontology. 49
2.11 Ontological Conitive Map. 49
2.12 Result of insering ontology model in a product synergy. . . 50

3.1 Example 1 of log-linear model. 52
3.2 Example 2 of log-linear model with two parents. 54
3.3 Example 3 of log-linear model with three parents. 55

A.1 Flow of information. 62

Liste des tableaux

1.1 The $\otimes$-and $\oplus$-operators for combining signs. 10

3.1 Links between semantics of QPN and $\mu - terms$ 58

Introduction

Decision making is a complicated task for decision makers, as they need to deal with knowledge from various stakeholders, sources, and research disciplines. One of the major obstacles to this task is that stakeholders may have different interests and will therefore have different perceptions of the problems being addressed. Their ideas can be contradictory and they may misunderstand the problems, which raises the questions of how to deal with conflicting stakeholder interests and how to use stakeholder knowledge in the decision-making process. Decision support systems may potentially help to bridge the gap between knowledge and decision making. Many different methods for representing and reasoning exist. Recently, there has been a growing interest in the use of graphical models to represent domain knowledge of decision-makers. One of these approaches that can be used to structure problems from the perspectives of specific groups or individuals is cognitive map (CM), This approach produces graphical diagrams consisting of nodes representing variables, and arrows representing causal relationships among these variables. However, CM is used only for representational purposes that is to say it is used for visualization task and doesn't allow an efficient reasoning because computer-supported analysis is possible only when the model is extended with mathematical functions and quantitative information. Qualitative probabilistic networks (QPNs) were designed to overcome CM's drawback.QPNs are abstractions of Bayesian networks, with conditional probability tables summarized by the signs of qualitative relationships between variables. There are powerful algorithms associated with this formalism used for causal reasoning. These algorithms are based on the idea of propagating and combining these signs and this is one of the major drawbacks of these qualitative abstractions. In the other hand, although causal Bayesian networks(CBN) are a well-established method for reasoning under uncertainty and making inferences, the elicitation of the structure and parameters of the network in complex domains can be a difficult and time-consuming task. Moreover, the notion of probability is often not perceived or understood by domain experts.

This work is mainly composed of two parts : the first part aims to study the similarities and differences between these three models and propose firstly and principally a process for incremental construction of CBN starting from an initial modeling CM through QPN, in order to take advantage of each model. To get there, we study the crossing-over between qualitative networks and probabilistic counterpart, establishe the main correspondences between their elements and develop an incremental approach in which we make use of the two formalisms in an Incrementally way. We introduce a method to build knowledge based systems incrementally for

reasoning and visualization. We use existing methods that we illustrate in the section of "Exixting bridges" enjoying their benefits and offering solutions to their limitations and drawbacks. We propose a global architecture that starts by specifying only causal statements creating an initial Cognitive Map using an elicitation process including several domain experts because it offers an easy way to express individual judgment, thinking or beliefs about a given problem. However, drawing inferences in cognitive maps, especially when the problem is complex, may not be an easy task. So, using the cognitive map model is not sufficient to analyze and understand the impact of the different variables of the problem. That's why we propose as a first step to provide clarifications of the cause-to-effect to build Qualitative Probabilistic Networks and then to add quantitative information as it is available and as time permits for more efficient reasoning because qualitative probabilistic network allows modeling interactions in terms of qualitative influences and synergies. However, qualitative probabilistic network often reveals tradeoffs : relationships that are ambiguous due to computing qualitative influences. To resolve such tradeoffs, we propose a second approach that incrementally enriches nodes that contribute to qualitative relationships by transforming influences to tables of probabilities. Then we combine these tables with observational data to build our causal Bayesian network as a well-established method for reasoning under uncertainty and making inferences. However, the notion of probability is often not perceived or understood by domain experts. The visualization task is therefore more tedious and time-consuming for decision makers to describe and capture knowledge in a more comprehensive manner. So that, we also propose to proceed inversely, that is to say, how to transform our enriched causal Bayesian Network to QPN for easier visualization. Then,we propose the integration of semantic knowledge contained in the domain ontology to anticipate causal discovery task and for more efficient visualization and construction results.

Namely, the goal is to move from a simple structure to a more complex structure for complex task in which computer-supported analysis and reasoning are possible. Enjoying the time that we can spend with the expert.That is to say the more we have time with the expert the more we can complicate the structure and provide a fairer and more accurate reasoning.

This process has two important advantages. First, rather than build a new network from scratch for each new problem, it is desirable to be able to store and reuse network fragments, just as programmers store and reuse code. While the structure of a qualitative network model specifies which variables influence one another may be applicable to multiple different situations, then we can add exact numerical probabilities to this network for each situation being evaluated. Second, in designing a probabilistic network starting from a qualitative one for an actual problem, can bridge the gap between the uncertainty provided by mathematical and numerical details offered by the Bayesian-network formalism and the qualitative abstraction offered by CMs and QPNs.

In the second part ; our overall goal is to fill the probabilities tables of our CBN from information provided by our QPN using the log-linear model. For this part we happened to find links between semantics of QPN and

coefficients of log-linear model. We are currently seeking a generalization of these links to find an automatic method to transform a QPN to a CBN. This task is difficult and requires more time that's why this method is included in the future research part of this thesis.
An outline of the remainder of the paper is as follows. Chapter 1 corresponds to the state of the art, where we discuss the definition of causal maps, semantics and the process of making inferences with causal Bayesian networks, definition and reasoning with the qualitative probabilistic network formalism, and the definition of ontology model. Then, we discuss a comparative study where we talk about the similarities and differences between these models. In the last part of this section we show the interactions that exist between each pair of models. In chapter 2, we propose a process for constructing a Bayesian network starting from a causal map. In chapter 3 we just introduce a method for probabilities tables estimation based on log-linar model, by detrmining links between parameters of causal Bayesian network and terms of this model. then, we focus on links between semantic of QPN and interaction terms of log-linear model.

Graphical knowledge-based systems

1

Sommaire

1.1 Cognitive map 6
1.1.1 Introduction 6
1.1.2 Graphical representation 7
1.1.3 Reasoning with a cognitive map 10
1.1.4 Construction method 10
1.1.5 Extension of cognitive maps 12
1.2 Causal Bayesian Networks 13
1.2.1 Introduction 13
1.2.2 Definitions 13
1.2.3 Reasoning with causal Bayesian network 15
1.2.4 Construction of causal Bayes Nets 20
1.3 Qualitative Probabilistic Networks 22
1.3.1 Introduction 22
1.3.2 Definition 22
1.3.3 Qualitative influence and synergies 22
1.3.4 The properties of symmetry, transitivity and composition 25
1.3.5 Intercausal influence 25
1.3.6 Reasoning with QPN 26
1.3.7 Extension of QPNs 28
1.4 Ontology 29
1.4.1 Introduction 29
1.4.2 Definition 29
1.4.3 Use of ontology 30
1.4.4 construction of an ontology 31
1.5 Comparative study 31
1.6 Existing bridges 32
1.6.1 Interaction between CM & Causal BNs 32
1.6.2 Interaction between CM & QPNs 33
1.6.3 Interaction between CM & Ontology 33
1.6.4 Interaction between BNs & QPNs 34
1.6.5 Interaction between BNs & Ontology 34

In this chapter we introduce two type of graphical models ; qualitative and probabilistic networks. Then we show how these models can be extended and even combined to release new models according to the needs of decision makers. Beside, we show that available ontologies provide high level knowledge for the domain under study where we can use graphical models for learning and knowledge visualization. The recourse to ontologies is due essentially to their ability to capture the semantics of domain expertise. Therefore, the semantical knowledge contained in the ontology can turn out of a big utility to improve causal discovery.

1.1 Cognitive map

1.1.1 Introduction

(Tolman 1948) was interested almost exclusively in animal psychology. It attributes the authorship of the term cognitive map. Tolman showed that the rat running in a maze is exposed to stimuli. In other words, it leads to a result of these stimuli to the responses which actually occur. This map of the environment gets established in the rat's brain. And it's this tentative map, indicating roots and paths and environmental relationships, which actually determines what responses, if any, the animal will finally release. He emphasized the flexible nature of this behavior by specifying the rat showed a great ability to adapt changes made by the researcher, including changing the starting position of the rat in the maze. Moreover, he concluded by proving that the concept of cognitive map could help to get a better understanding of some psychological mechanisms. Therefore, a cognitive map can be used to refer to a representation of space stored in the brain. He emphasized that the map determinates behavior of the animals. Tolman has not provided concrete representations of a cognitive map, whereas he supported the idea of the human brain ,which contains maps dedicated to other tasks as the location in space and that these maps determinate our behavior.

Over the years the metaphor of the cognitive map is found at the heart of several studies, especially some works like (Axelrod 1976) and his team, who use cognitive maps to represent graphically individual's beliefs about how a given concept is responsible for the perturbation on the state of another concept.

Furthermore, cognitive map describes a problem and helps to predict individual behavior that faced this problem. (Axelrod 1976) proposed a mechanism of inference. The goal is to capture influence relations, called bonds, between different notions : a notion is either favorable or not favorable to another notion. This kind of judgment is known as cause-effect relationships between concepts. This mechanism helps to evaluate different alternatives, discover unanticipated effects and design scenarios to solve the problem. Since then a broad spectrum of application has been carried out in many fields likewise, decision support (Kosko 1986), Probabilistic causality (Wellman 1990), virtual reality (Dickerson et Kosko 1994), Multi-Agent System (Chaib-draa 2002), On-Line Analytical Process (Zhang et al.

2003), and much more. Figure 1.1 shows schematically the evolution of work on cognitive maps.

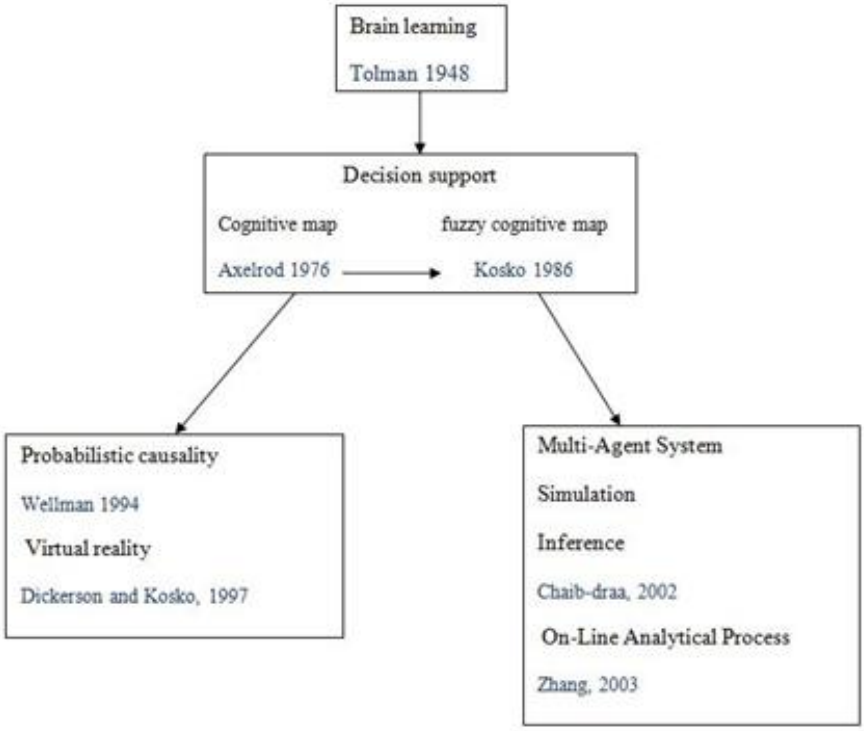


FIGURE 1.1 – *The evolution of work on cognitive maps.*

1.1.2 Graphical representation

Concepts and links

(Chauvin et al. 2011) defined a cognitive map as follow :

Definition 1.1 *(Cognitive map). Let C be a concept set and I a value set. A cognitive map CM defined on I is a directed graph CM = (C,A, $etiq_I$) where :*
- *the concepts of C are the nodes of the graph ;*
- *$A \subseteq C \times C$ is a set of arcs, called in uences ;*
- *$etiq_I : A \rightarrow I$ is a function labelling each influence with a value of I.*

A cognitive map contains two kinds of information : on the one hand, nodes called states represent notions or concepts, the label of a vertex can be interpreted in many angles, it can represent both the presence of the concept or its level. For instance, the vertex labeled as benefit might be interpreted as the presence of benefit or as an amount of profit. It should be noted that a map can not contain two vertices with the same label. A label describes a concept but also implicitly its opposite. The contrary appears implicit in the case where concepts are connected by a link of negative polarity. Sometimes notions are treated as variables that is to say, as factors may take different values as in the work of (Axelrod 1976), (Bougon et al. 1977) and many others. They may be dichotomous (e.g. : the presence or absence of a union in the company), ordinal (e.g. : the level of satisfaction among employees), rational scale (e.g. : the number of years of

existence of the company) and others. On the other hand, cognitive map contains directed edges between those nodes represent influence links, called also bonds -generally positive or negative- between some of these concepts.

Links between concepts are important because they define the structure of the card. (Axelrod 1976) defines links of a cognitive map as influences' links or form influences' trails that is to say paths connecting one concept to another by passing through one or more other concepts. This concept of influence is unclear ; (Hume 1739) prefer to define links as causal link . Smoking is the cause of cancer is an example of a causal link between the concept of smoking and the concept of lung cancer. Cognitive maps containing causal links are called causal maps (Bougon et al. 1977). A causal relationship is not equivalent to a logical implication (Kosko 1986) because the opposite is not verified. Notably, do not have lung cancer because we do not smoke is not true, however, non-smoking tends not to cause lung cancer. Causal link is a special case of influence link that should express a mechanical relationship, e.g. performance leads growth (Allard-Poesi 1996).

To understand the remarks above we use Figure 1.2. In this figure a plus sign indicates the existence of a positive relationship, in other words , a change in the level of a variable leads to a change in the same direction of variable's level to which the link is oriented. A negative sign indicates a negative relationship which means that the variation of a variable leads to a contrary variation of the variable to which the link is oriented.

The indirect effect of an initial variable on final variable is positive if the path doesn't contain or if it contains a pair number of negative sign, as shown in Figure 1 the trail AED higher expenses advertising lead to higher business performance via increasing sales. In addition, the indirect effect of one variable on another will be negative if the number of minus signs is odd. This is the case of ACD path that an increase in advertising spending would lead to a decline in the performance of the company because it would increase costs.

To investigate the use of a cognitive map we must obviously learn the characteristics of the links that can compose it and types of questions that can be asked and for what we look for answers via a map's reading. A link can be characterized by a set of characteristics. We have identified the following attributes : polarity, power (force), probability, and latency. In what follows we present these characteristics by illustrating them

Polarity

Most models of cognitive maps associate polarities to links. A link is either positive or negative. The polarity of a link can be determined by asking the following question : Does the cause have a positive or negative influence on the effect ? Considering the cognitive map represented in figure 2,the relations between concepts are represented by polarities. The cognitive map model also defines an inference mechanism that consists into the propagation of influences. This inference process is the basis of the decision support process.

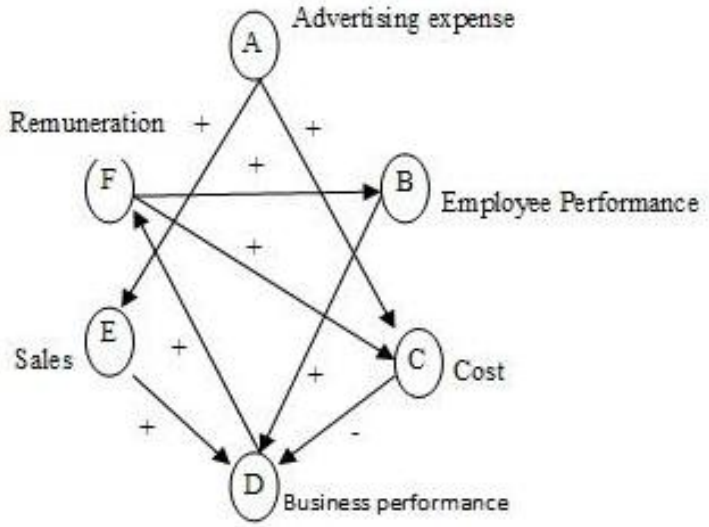


FIGURE 1.2 – *Example of Cognitive Map.*

Force

The strength of a relationship between two concepts describes proportion xhere the concept-effect evolves according to the concept-cause. The force of a relationship can be determined by asking the following question : To what extent the cause causes an increase in the effect ? Force is represented by an element belonging to an ordered set. For example, some models of cognitive map represent the force as a real number between 0 and 1, other by an integer between 1 and 5, or a symbolic value of the set low, medium, and lot. For example, a positive link between the concept smoking and the concept cancer is interpreted as that smoking has a positive influence on cancer.
Considering the concepts advertising expense and sales as variables, increase the amount of the first factor increases the possibility of getting lot of sales.

$$\boxed{\text{Advertising expense}} \xrightarrow{+\text{lot}} \boxed{\text{Sales}}$$

Probability

Some models characterize a link by a probability value. The probability value of a link can be determined by asking the following question : What is the probability that the effect occurs if the cause occurs ? For example, a positive probability of 0.66 between the concept advertising expense and the concept sales mean that two out of three, the advertising expense increases the probability of sales

$$\boxed{\text{Advertising expense}} \xrightarrow{+0.66} \boxed{\text{Sales}}$$

Latency time

The latency of a link can be determined by asking the following question : What is the time required for the effect to occur once the cause has occurred ?

This value can be measured by expressing duration in minutes, hours, etc, or using a symbolic scale "immediate, short term, medium term, long term, very long term".

1.1.3 Reasoning with a cognitive map

Reasoning with a cognitive map describes the effect produced by a concept to another based on paths that exist in the map between these two concepts, and labels carried by the links that make up these paths. this effect can be positive (+), negative (-), zero (0), or ambiguous (?). Both operations are used to measure the influence of a concept to another : the addition, which corresponds to the idea of cumulative influences of several paths having same destination state and multiplication corresponds to the idea of transitivity influences in a path. (Genest et Loiseau 2007) defined the inference mechanism as follows :

Definition 1.2 *(influence) the influence I of X is a $C \times C$ $\{+, -, 0, ?\}$ such that :*

$$I(e_i, e_j) = \oplus_{H \in H_{i,j}} \otimes_{i \in [1,|H|-1]} I_1(h_i, h_{i+1}) \tag{1.1}$$

$H_{i,j}$ is the set of paths H that have th origine e_i and e_j extremity. H a path composed of arcs ($e_1 = h_1$, h_2) (h_2, h_3) ... (h_{n-1}, $h_n = e_j$) and $|H|$ is the length of the path.
I_1 is an application of $C \times C$ in $\{+, _0\}$ such that $I_1(e_i, e_j) = tiq_X(l)$ if there is an $l = (e_i, e_j)$ in L_X and 0 otherwise.
$\oplus$ and $\otimes$ are applications $\{+, _0, ?\} \times \{+, _0, ?\}$ In $\{+, _0, ?\}$ defined as follow :

$\otimes$	+	-	0	?
+	+	-	0	?
-	-	+	0	?
0	0	0	0	0
?	?	?	0	?

$\oplus$	+	-	0	?
+	+	?	+	?
-	?	-	-	?
0	+	-	0	0
?	?	?	?	?

TABLE 1.1 – *The $\otimes$-and $\oplus$-operators for combining signs.*

1.1.4 Construction method

To build a cognitive map, we have to go through these two important phases proposed by (Axelrod 1976) :

- Information's source selection.
- Acquisition of the cognitive map.

Information's source selection

There are two sources of information : documents and experts. Before the creation of a cognitive map it is better to check if you have chosen

the right document, that is to say, if the document contains clear ideas to achieve this goal there are four requirements that the document should verify :

- The document must contain sufficient causal assertions, to provide material for analysis.
- The document must have sufficient information on the field so that the encoder can understand the meaning of words.
- The document must be complete, that is to say, it contains all the necessary and relevant information on the domain.
- The document must be a reliable representation of reality.

For the method using a panel of experts to create a cognitive map is a good choice. In this case, the cognitive map is usually more accurate than a map built with the opinion of one expert.

The acquisition of the cognitive map

(Axelrod 1976) has proposed to acquire a map either from a document or from a group of experts (by questionnaire).

Acquisition from a document **:** Once the correct document is selected, a general methodology is to go through the document and identify influential links in the form of triplets cause / link / effect. The similar triplets are merged. Concepts representing similar causes or similar effects are also merged.

Then we make two lists, one for the concepts and other for causal assertions, where both lists are extracted from the document and reflect the relationships of cause and effect between concepts.

It is not necessary to keep the original word that was in the text, that is to say, names of concepts may be similar but not necessarily identical. The set of concepts and links obtained are finally presented in graphical form and the concepts must be connected by causal relationships to build the map.

Acquisition by a group of experts **:** You identify relevant concepts, and we can ask experts to write a list of all relevant concepts related to the field that we hope analyze. We can limit the number of concepts in assessing their importance. Then, we choose the weight and arcs, in this step we have to distribute questionnaires to the experts with all possible combinations of concepts previously selected. The expert must assign a sign of -, 0, + for each pair of different concepts. Then we have to collect the map of each expert and integrate them into a single cognitive map. Once the resulting map is finished, we can now return it to experts for discussion and change. If there are disagreements, they should be identified and analyzed. The map is usually amended several times before the experts are satisfied with the final result. there is a method that is based on this idea called self-Q (Bougon 1983), but in this method those who ask questions are experts themselves. That is to say, participants are experts on the personal knowledge that guides their social behaviour. They formulate their questions on the basis of their own personal knowledge.

1.1.5 Extension of cognitive maps

Fuzzy Cognitive Maps have emerged as alternative tools for representing and studying the behavior of systems and people. FCM was developed by (Kosko 1986). They were created for representing causal reasoning. Their fuzziness allows hazy degrees of causality between hazy causal concepts. Causality as represented as a fuzzy relation on causal concepts, that is to say values in the graph are fuzzy, so concepts take values in the range between $[0, 1]$ and the weights of the arcs are in the interval $[-1, 1]$. Weights of arcs between concept C_i and concept C_j could be positive $(W_{ij} > 0)$ or negative $(W_{ij} < 0)$ and the theory behind that computation is fuzzy logic.

FCM facilitates learning and improves the inference mechanisms developed by (Axelrod 1976). It is a tool for analyzing complex phenomena such as the factors influencing the physiological capacities of diabetic patients (Taber 1991), the analysis of electric circuits (Styblinski et Meyer 1988), factors influencing the viscosity of fluids (Banini et Bearman 1998), a simulation tool (Miao et al. 1999), (Park et Kim 1995), (Carvalho et Tomé 1999), a tool for controlling the agents in a virtual world (**?**), (Chaib-draa 2002), (Mata Avila 2002), (Truck 2002). Inference mechanisms are redesigned to see the evolution over time of values associated with the concepts of the map. We can provide an example of FCM,Figure 1.3, using the example already illustrated previously were "+" and "-" signs are replaced by scalar values between [-1,1] that are used to compute the "strength of impact" of a concept on an other one.

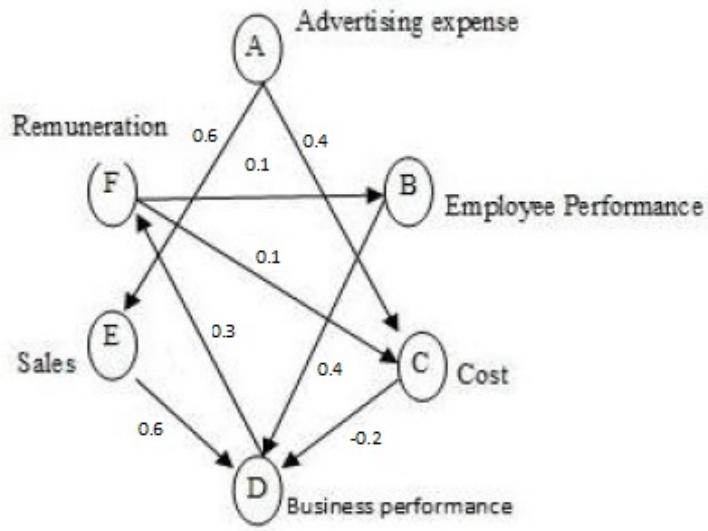


FIGURE 1.3 – *Example of Fuzzy Cognitive Map.*

1.2 Causal Bayesian Networks

1.2.1 Introduction

The Bayesian networks have emerged in the first half of the 1980s in artificial intelligence. More precisely, they were introduced in this field as a tool for dealing with uncertainty.

A Bayesian network or simply Bayes net introduced by (Judea 2000) is the Couple consisting of a directed acyclic graph and a probability distribution ; it represents the probability distribution of a set of random variables in combination with dependent variables using graphs. It is so that the combination of the two science graph theoreis on the one hand and the other probabilistic notions.

These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. Indeed, nodes that are not connected by edges are variables that are conditionally independent of each other and nodes that are connected by edges do have conditional dependent relationships. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Causal Bayesian Network is a particular type of Bayesian Network where egde directions indicate direct causality. Recently, Causal Bayesian Network become very used because causal relationships have more "semantics" than dependence relationships. However, its construction is more difficult because it needs observational and interventional data.

1.2.2 Definitions

Bayesian network

Qualitatively, Bayesian network is known as a directed acyclic graph (DAG) and quantitatively it enables a representation and computation of the joint probability distribution (JPD) over a set of random variables. This JPD can be determined uniquely by these local conditional probability tables (CPTs). Formally the network is defined by a pair $B =< G, \theta >$ where G is the DAG whose nodes (vertices) $X_1, X_2, \ldots, X_n$ represents random variables or roughly speaking that variable X_i influences X_j. Node X_i is then referred to as a parent of X_j and, similarly, X_j is referred to as the child of X_i . And directed edges represent the direct dependencies between these variables. A BN reflects a simple conditional independence statement. Namely that each variable is independent of its non descendents in the graph given the state of its parents. This property is used to reduce, sometimes significantly, the number of parameters. So to build a Bayes net first it must identify the variables of interest. This involves answering the question : what are the nodes to represent and what values can they take, or what state can they be in ? Then to define the structure, we might ask what factors affect another factor then we should add arcs between them.

The second component denotes the set of parameters of the network ;

$\theta = \{P(X_i|\pi_i)\}$: represents the set of probabilities for each node X_i conditional on the state of its parents π_i in G. Accordingly, B defines a unique JPD over V, namely :

$$P_B(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_B(X_i|\pi_i) = \prod_{i=1}^{n} \theta_{X_i|\pi_i} \tag{1.2}$$

This function is decomposed into a product of terms depending only on the local node and its parents in the graph that is a very important result and very useful of fundamental property (independence assumption) already mentioned above that verifies each variable Xi is independent of its non descendents given its parents in G. This is an example of Bayesian network introduced by (Judea 2000) modeling the problem of lung cancer A patient has been suffering from shortness of breath (called dyspnoea) and visits the doctor, worried that he has lung cancer. The doctor knows that other diseases, such as tuberculosis and bronchitis, are possible causes, as well as lung cancer. She also knows that other relevant information includes whether or not the patient is a smoker (increasing the chances of cancer and bronchitis) and what sort of air pollution he has been exposed to. A positive X-ray would indicate either TB or lung cancer. In Pearl's medical diagnosis example, we might ask what factors affect a patient's chance of having cancer ? If the answer is "Pollution and smoking," then we should add arcs from Pollution and Smoker to Cancer. Similarly, having cancer will affect the patient's breathing and the chances of having a positive X-ray result. So we add arcs from Cancer to Dyspnoea and X-Ray. The resultant structure is shown in Figure 1.4.

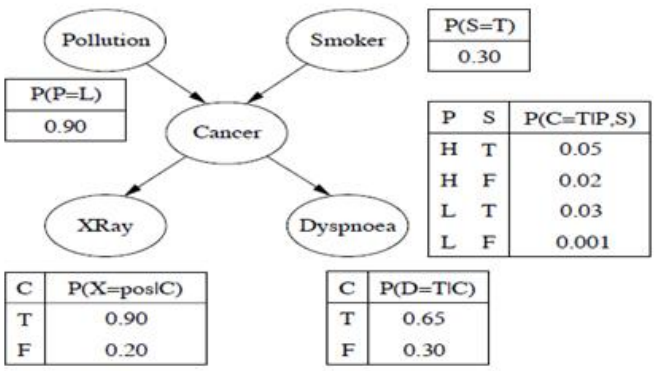


FIGURE 1.4 – *Example of Bayes Networks.*

Causal Bayesian network

Causal Bayes nets have been developed in the philosophy of science and statistical literature over the last fifteen years (Judea 2000), (Spirtes et al. 2001) scientists seem to infer theories about the causal structure of the world from patterns of evidence.

The basic idea of causal Bayesien network that arrows represent the Bayesian relationship between direct cause and effect variables on which all the network is defined. To put it more concisely, a causal Bayesian network is a Bayesian network whose graph is causal.

The advantages of building DAG model around causal that is the judgments required in the construction of the model are more meaningful more accessible and hence more reliable and the ability to represent and respond to external or spontaneous changes because Bayes nets are based on a very fundamental notion which is the intervention that is to say the development of causal Bayes net algorithms allow us to determine what will happen when we intervene from outside to change the value of particular variable when two variables are genuinely related in a causal way, holding other variables constant, the intervening to change one variable should change the other. Predictions about probabilities may be quite different from predictions about interventions.
The intervention assumption : A variable I is an intervention on a variable X in a causal graph if and only if I is exogenous (that is it is not caused by any other variables in the graph), directly fixes the value of X to x and does not affect the values of any other variable in the graph except through its influence on X.

1.2.3 Reasoning with causal Bayesian network

Reasoning with causal Bayesian Network called "inference". Scientists infer causal structure by observing the patterns of conditional probability among events by examining the consequences of interventions.

The use of causal Bayes nets has an important objective. Knowing the causal structure let us make the right predictions about interventions and probabilities. In causal Bayes nets interventions alter the nature of the graph they intervene on.

Causal structures help us to answer to many questions such as what would have happened, had things been otherwise ? By fixing the value of certain variables and inferring the consequences.

Reasoning with causal Bayesian networks is different from that with BNs because there are intervention data that we have to take into cisideration. However, in BNs we make inference just from observations that's why we must distinguish between the notion intervention and observation concept then we will show the method of mutilation-augmentation to manipulate

Observation vs. intervention

The formal distinction between the two notions is an important prelude to the rest of this thesis :
By referring to LAROUSSE dictionary, intervention is an action, role, an effect of something in a process. Otherwise and in relation with causal Bayesian networks intervention is the fact to act on a variable and see its impact or influence on other variables. The same dictionary defines observation as the act of watching carefully phenomena, events, things to study, monitor, draw conclusions, etc.
Normatively, predictions based on observed values of a variable often differ from inferences drawn from the very same states brought about by active interventions.

(Judea 2000) propsed to write interventions with a "do-operator", written as do (.).
Whereas the expression $P(a|c)$ refers to the probability of $A = a$, given that $C = c$ was observed, the expression $P(a|do(c))$ refers to the probability of $A = a$, given that C is fixed to state c by means of an intervention.
Formally (Woodward 2003), (Judea 2000), (Spirtes et al. 2001), and Scheines defined intervention as follow :

Definition 1.3 *(Intervention) Given a set of observed variables V, an intervention I on a subset $S \subseteq V$ must satisfy the following conditions :*

- *$I \notin V$ is a variable with two possible states (on/off) representing where the intervention can be active or inactive.*
- *I directly manipulates each variable $X \in S$,*
- *I is exogenous to V*
- *When $I = off$, the joint distribution over V obtains, i.e.*

$$P(V|I = off) = \prod_{V_i \in V} P(V_i|pa(V_i)) = P(S|pa(S)) \prod_{V_i \in V\ S} P(V_i|pa(V_i)) \tag{1.3}$$

- *When $I = on$, the conditional distribution over S is manipulated , i.e.*

$$P(V|I = on) = P(S|pa(S), I = on) \prod_{V_i \in V\ S} P(V_i|pa(V_i)) \tag{1.4}$$

where

$$P(S|pa(S), I = on) = \prod_{X \in S} P^*(X|pa(X)) \tag{1.5}$$

and for each $X \in S$, we have :

$$P^*(X|pa(X)) \neq P(X|pa(X), I = off) \tag{1.6}$$

Pearl defined an exogenous variable as a factor in a causal model or causal system whose value is independent from the states of other variables in the system ; a factor whose value is determined by factors or variables outside the causal system under study.

There two kinds of interventions ; structural and parametric interventions. The first kind has an influence on the network structure. In fact if we add an exogenous variable such that its state is on and a single arrow into the manipulated variables of the graph then the entire edges incident on any intervened variable are removed. This change in causal structure implies a change in the joint probability distribution over the variables. And the altered graph will be called post-manipulation graph. The second kind of intervention called parametric interventions and it's a weaker form of intervention we use it when it is not necessary to change the structure of network.

For more understanding the seeing and doing fact we use formulas probability. Conditional probability that variable $Y = y$ when we see that $X = x$ is noted :

$$P(Y = y|see(X = x)) = P(Y = y|X = x) = P(y|x) \tag{1.7}$$

Conditional probability that variable $Y = y$ when we set X to x is noted :

$$P(Y = y|do(X = x)) = P(y|do(x)) \tag{1.8}$$

If Y is the cause of X then $P(y|do(x))$ is equal to $P(y)$
If Y is the direct effect of intervening on X then $P(y|do(x))$ is equal to $P(y|x)$

Intervention as a negation of all other direct causes :

In graphical models, interventions on a variable A_i also express our beliefs about parents U_i of A_i. This is shown by deletion (also called mutilation) of links between U_i and A_i. The rest of the graph remains intact. This graph model this situation where beliefs about A and B do not change.

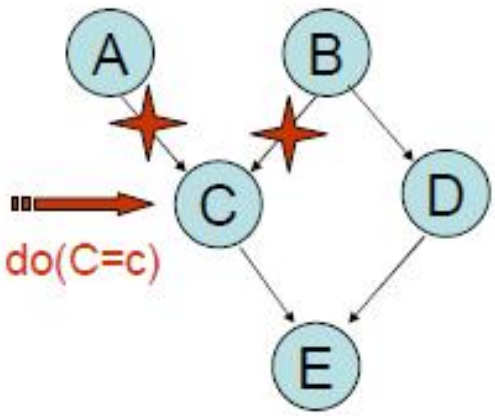


FIGURE 1.5 – *An altered graph after mutilation*

Using the graph, all other direct causes (parents) other than the action becomes independent of the variable of interest. In addition, the event which combines the variable A_i has the value after completion of the intervention do (a_i) becomes certain. Formally, The effect of such intervention is given as follows :

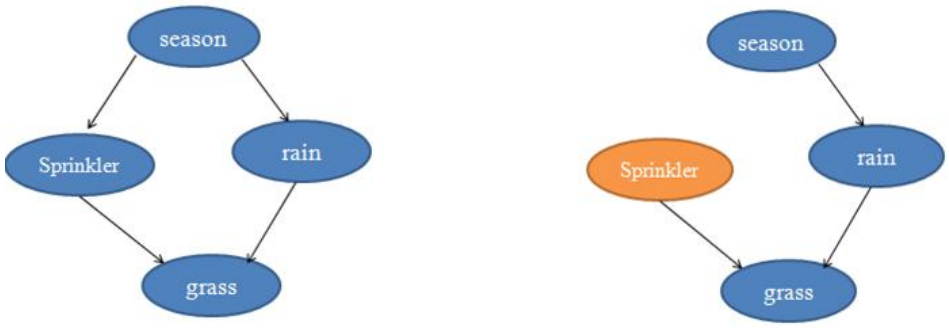
$$P(a_1, \ldots, a_i, \ldots, a_n | do(a_i^{'})) = \begin{cases} \prod_{j \neq i} P(a_j | u_{A_j}) & \text{if} \quad a_i = a_i^{'}, \\ 0 & \text{if} \quad a_i \neq a_i^{'}. \end{cases}$$

Example : Figure 1.6 shows how the probability distribution is changed after we intervene on the variable "Sprinkler". We note that the arc between "Season" and "Sprinkler" was deleted

Adding additional nodes :

An alternative but equivalent approach proposed by (Judea 1993) is to consider the intervention as a new variable in the system. The effect of an intervention is calculated by applying the conditioning on an altered graph (increased). This alteration is to add a parent node DO_i to each variable A_i that may be affected by an intervention. The resulting graph is said to be increased.

Do_{A_i} is a variable that takes value in $\{do(A = a), No\}$ for all instance a' of A, and "No" means "no action" or (nothing). The new local probability



$P(SPGR) = P(S)P(P|S)P(R|S)P(G|SR)$ $\quad$ $P(SPGR) = P(S)P(R|S)P(G|SR)$

FIGURE 1.6 – *Example of CBN before and after mutilation*

distribution at the variable A_i after increasing the graph is given by :

$$P(a|u_{A'}) = \begin{cases} P(a|u_A), & \text{if} \quad DO_A = No \\ 1, & \text{if} \quad DO_A = do(a') \quad \text{and} \quad a' = a, \\ 0, & \text{if} \quad DO_A = do(a') \quad \text{and} \quad a' \neq a. \end{cases}$$

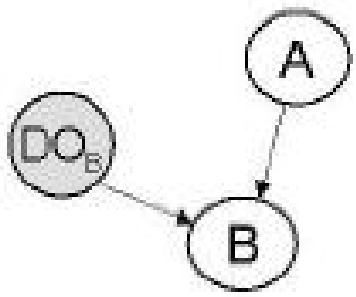
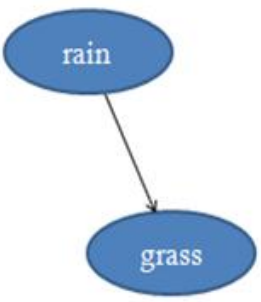


FIGURE 1.7 – *The augmented graph G_a.*

Example

Rain	
yes	0,8
no	0,2

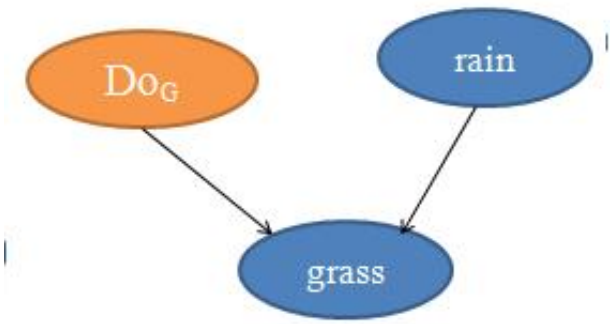


Grass	**Rain**	**Grass | Rain**
wet	yes	0,9
wet	no	0,3
dry	yes	0,1
dry	no	0,7

If there is intervention Do(grass=wet) :

nothing	0
do(wet)	1
do(dry)	0

Rain	
yes	0,8
no	0,2



Grass	Rain	DO_G	P(Grass | Rain,DO_G)
wet	yes	nothing	0
wet	yes	do(wet)	0,8
wet	yes	do(dry)	0
wet	no	nothing	0
wet	no	do(wet)	0,2
wet	no	do(dry)	0
wet	no	nothing	0
dry	yes	do(wet)	0
dry	yes	do(dry)	0
dry	no	nothing	0
dry	no	do(wet)	0
dry	no	do(dry)	0

If there no action (*nothing*) :

Grass	Rain	DO_G	P(Grass | Rain,DO_G)
wet	yes	nothing	0,9
wet	yes	do(wet)	1
wet	oui	do(dry)	0
wet	no	nothing	0,3
wet	no	do(wet)	1
wet	no	do(dry)	0
wet	yes	nothing	0,1
dry	yes	do(wet)	0
dry	yes	do(dry)	1
dry	no	nothing	0,7
dry	no	do(wet)	0
dry	no	do(dry)	1

After adding or deleting arcs we can use the same propagation algorithms that we use with classic BNs such as message passing, junction, etc.

1.2.4 Construction of causal Bayes Nets

This step corresponds to the structure learning phase. In this phase we can use the methods based on conditional independence tests such as PC (Spirtes et al. 2001) and PowerConstructor [Cheng et al., 2001]. These algorithms are divided into three phases :

1. Search conditional independencies in the data using statistical tests.
2. Using independence discoveries to form a partially directed graph without circuit (PDAG) removing edges completely connected to each independent discovered ($X \rightarrow Y$). The undirected graph obtained will be then partially directed through the conditional independencies discovered with keeping V-structure $X \rightarrow Y \rightarrow Z$.
3. The PDAG obtained is then "completed" (CPDAG) with edges so you do not create a v-structure or cycle these edges are called "inferred edges" In the case of causal Bayesian networks, the choice of DAG in the equivalence class represented by the previous CPDAG can not be performed in the same manner. The choice of DAG can not be arbitrarily. We can not only use the observational data used in learning CPDAG because we will obtain a graph in witch there is some undirected arcs. These types of links don't allow us to know witch is the cause node and witch is the effect one. we distinguish between two types of methods. Methods based in both observational and experimental data in witch they add a series of experiments. The basic idea of these methods that if we found two linked variable we can not decide which variable is the cause of which variable in this case we need to manipulate a variable and then observe the potential impact on other variables to find the correct causal structure For example, if we handle X and we observe a change in Y, we know that the causal structure is $X \rightarrow Y$. Moreover, the order in which these experiments are performed can affect the speed of the learning algorithm. The second kind of method is based only on observational data.

(Cooper et Dietterich 1992) propose score based methods for learning Bayesian networks from a combination of prior knowledge and statistical data. These approaches were revisited by [Cooper and Yoo., 1999] to be suitable for learning causal BNs by combining an arbitrary mixture of observational and experimental data.

(Tong et Koller 2001) developed an active learning method to guide the interventions to collect new interventional data for further structure probability updates. They formalize the causal learning task as that of learning the structure of a causal Bayesian network. They consider an active learner that is allowed to conduct experiments, where it intervenes in the domain by setting the values of certain variables. This method is based on the current belief about the structure and the causal information that will be gained by an experiment. (Murphy 2001) proposed a similar technique where different approximations are used to overcome working in the space of DAGs. Murphy used MCMC to approximate the belief state and importance sampling to calculate the expected utility.

(Eberhardt et al. 2005) proved that, under ideal conditions with causal Markov assumption and faithfulness assumption (and ideal distributions), the number of interventions required to identify the causal relationships among N variables is N -1 when only one variable can be manipulated each time, and the number of interventions is N log2 when multiple variables can be manipulated simultaneously.

(Meganck et al. 2006) assumed that the correct complete partial directed acyclic graphs (CPDAGs) can be learned from the observational data and the directions of the undirected edges in the CPDAGs can be determined with interventional data. They proposed a greedy approach known as MYCADO (My Causal Discovery) algorithm based on maximax and maximin decision criterions and where the choice of best experiment depends on calculating a utility function U(AX_i).

(Borchani et al. 2007) proposed a new method for learning causal Bayesian networks from incomplete observational data and interventions. They extended the Greedy Equivalence Search-Expectation Maximization (GES-EM) algorithm initially proposed to learn Bayesian networks from incomplete observational data, by adding a new step allowing the discovery of correct causal relationships using interventional data. Two intervention selection approaches are proposed : an adaptive one, where interventions are done sequentially and where the impact of each intervention is considered before starting the next one, and a non-adaptive one, where the interventions are executed simultaneously.

(He et Geng 2008) developed a framework for active learning of causal structures via intervention experiments for discovering causal structures in which they first find a Markov equivalence class from observational data, and then they orient undirected edges in every chain component via intervention experiments separately. In the experiments, some variables are manipulated through external interventions. They discussed two kinds of external intervention experiments : the randomized experiment and the quasi-experiment. In order to reduce the complexity of the causal discovery task, the authors proceed by splitting the Markov equivalence class into subclasses and making experimentations on chain components. They also proposed optimal designs of batch and sequential interventions based on the minimax and the maximum entropy criteria.

Recently, (Ben Messaoud et al. 2009) introduce serendipitous decision-making strategy, referred to by SEMCADO (Semantical Causal Discovery), it's an extension of MYCADO algorithm by integrating ontological knowledge for more efficient causal discovery. They also introduced the notion of serendipity which is very useful when setting the experimental design. The idea mainly arises from the fact that the semantically related concepts are generally the most extensively studied ones.

1.3 Qualitative Probabilistic Networks

1.3.1 Introduction

Qualitative Probabilistic Networks (QPN for short) are basically qualitative derivations of Bayesian belief networks. Originally, QPNs were designed to improve the speed of the construction and calculation of these networks, at the cost of specificity of the result. The formalism can also be used to facilitate cognitive mapping by means of inference in sign-based causal diagrams. QPN was introduced by (Wellman 1990) and it is a network graph model with nodes that represent variables and signed edges and hyper-edges describe relationships among them.

1.3.2 Definition

Formally, a QPN consists of an acyclic digraph G = (V(G) ;A(G)) where V is the set of nodes in the graph representing the variables characterizing the problem domain under consideration. The set of arcs A(G) models the independences holding between the variables, QPN are influence diagrams where the numeric conditional probability tables are replaced by qualitative probabilistic relations among the variables. This model is the combination between qualitative influences and synergies represented in the graph by direct edges and hyper-edges respectively in the graph.
To define a QPN we need to distinguish between three types of qualitative probabilistic relationship : qualitative influences, additive synergies and product synergies. We will now give the definition of each type of influence, which is derived from (Wellman 1990).

1.3.3 Qualitative influence and synergies

Qualitative influence

All qualitative relations take the form of conditions on the comparative probability distribution for a variable given various values for its predecessors in the networks. So each variable has a cumulative conditional probability distribution function as a function of its predecessors. In QPN, a qualitative relation contains the value of this function without specifying it precisely. This type called qualitative influences, restrict the relative values of a cumulative conditional probability distribution function of a variable upon variation of one of its predecessors.

Definition 1.4 *Let U be a set of variables. Let Pr be a joint probability distribution on U and let $B \in U$ with values $b_1 < \cdots < b_n, n \geq 1$. Then, the function $F_B : \{b_1, \ldots, b_n\} \longrightarrow [0, 1]$ defined by*

$$F_B(b_i) = Pr(b_1 \vee b_2 \vee \cdots \vee b_i)$$

is the cumulative probability distribution fonction of B.

Consider two variables, A and B, when A and B denote events, a qualitative influence is a statement of the form A makes B more (or less) likely. So higher values for a node B are more likely given higher values for a node A, if the cumulative conditional probability distribution $F_{B|a_i}$ of

node B given a_i lies, graphically speaking, below the cumulative conditional probability distribution $F_{B|a_j}$ given a_j for all values a_i, a_j with $a_i > a_j$. $F_{B|a_i}$ is said to dominate $F_{B|a_j}$ by first-order stochastic dominance (FSD)

Definition 1.5

$$F'_{B|a_i}\, FSD\, F_{B|a_j} \Leftrightarrow F'_{B|a_i}(b_i) \leq F_{B|a_j}(b_i) \textit{ for all values } b_i \textit{ of } B$$

Example

We consider two binary nodes A with values values a_i and a_j , $a_i > a_j$ and B with $b_1 < b_2 < b_3 < b_4$. The following conditional probability distributions are specified for B, given A

$$\begin{array}{ll} Pr(b_1|a_i) = 0.25 & Pr(b_1|a_j) = 0.40 \\ Pr(b_2|a_i) = 0.35 & Pr(b_2|a_j) = 0.40 \\ Pr(b_3|a_i) = 0.25 & Pr(b_3|a_j) = 0.20 \\ Pr(b_4|a_i) = 0.15 & Pr(b_4|a_j) = 0.00 \end{array}$$

The cumulative conditional probability distributions for node B given A then are :

$$\begin{array}{ll} Pr(B \leq b_1|a_i) = 0.25 & Pr(B \leq b_1|a_j) = 0.40 \\ Pr(B \leq b_2|a_i) = 0.60 & Pr(B \leq b_2|a_j) = 0.80 \\ Pr(B \leq b_3|a_i) = 0.85 & Pr(B \leq b_3|a_j) = 1.00 \\ Pr(B \leq b_4|a_i) = 1.00 & Pr(B \leq b_4|a_j) = 1.00 \end{array}$$

A qualitative influence from A to B with sign $+(-)$ means that the probability distribution for B given A in nondecreasing (nonincreasing) in A in the sense of first order stochastic dominance. The notation $S^{\delta}(a,c), \delta \in \{+,-,0,?\}$ is used to denote such an influence.

Let $G = (V(G); A(G))$ be an acyclic digraph and let Pr be a joint probability distribution on $V(G)$ such that G is an I-map for Pr. Let $A, B, C \in V(G)$ be nodes in G with $A \to C$, $B \to C \in A(G)$.

Definition 1.6 ***(qualitative influences)*** *$S^{\delta}(A,B)$ holds in qualitative probabilistic network G in and only if (iff), for all values b_i of B, x of B's predecessors in G other than A, and $a_i > a_j$ of A :*

$$P_B(b_i|a_i x)\, R_{\delta}\, P_B(b_i|a_i x),$$

where R_{δ} is $\leq, \geq$, or $=$ as δ is $+, -$, or 0, and $R_?$ is the complete relation (thus $S^?$ always holds).

Consider binary variables. This binary case is easy to capture in a probabilistic assertion. Let a and $\bar{a}$ denote the assertions A= true and A= false, respectively, such that $\bar{a} \geq$a. For binary nodes the definition of qualitative influence can be simplified.

Definition 1.7 *Let G, Pr, A, B and X be as in the previous definition. Then, node A positively influences node B along arc $A \to B$ in G iff*

$$Pr(b|ax) \geq Pr(b|\bar{a}x),$$

for any combination of values x for the set of nodes X.

Additive synergy

An additive synergies relation is intended to capture the intuition that an increase in two variables together has greater effect at higher level of another. In other words, the combined effect of increasing the variables is greater than taking the two effects independently. To better explain, we can take an example of a positive additive synergy of nodes A and B on a common child C expresses that the joint influence of A and B on C is greater than the sum of their separate influences, regardless of any other direct influences on C.

Let $G = (V(G); A(G))$ be an acyclic digraph and let F be a joint probability distribution on $V(G)$ such that G is an I-map for Pr. Let $A, B, C \in V(G)$ be nodes in G with $A \to C$, $B \to C \in A(G)$. and for simplicity we will take binary nodes A, B, and C.

Definition 1.8 ***(Additive synergies)*** *$Y^{\delta}(\{A, B\}, c)$ holds in QPN G iff for a value c_0 of chance node c, x of c's prodecessors in G other than a and b, $a > \bar{a}$, and $b > \bar{b}$ of b :*

$$[F_c(c|abx) + F_c(c|\bar{a}\bar{b}x)]\ R_{\delta}\ [F_c(c|a\bar{b}x) + F_c(c|\bar{a}bx)].$$

where R_{δ} is $\leq, \geq$, or = as δ is +, -, or o, and $R_?$ is the complete relation (thus $S^?$ always holds).

Example

$$Pr(w|ul) + Pr(w|\bar{u}\bar{l}) = 0.30, \quad \text{and} \quad Pr(w|\bar{u}l) + Pr(w|u\bar{l}) = 0.28.$$

$$Pr(w|ul) + Pr(w|\bar{u}\bar{l}) \geq Pr(w|\bar{u}l) + Pr(w|u\bar{l}),$$

Then nodes U and L exhibit a positive additive synergy on node W, that is, $Y^{+}_{G}(\{U, L\}, W)$.

Product synergy

A product synergy is intended to capture the intuition that an increase in one variable has greater effect at higher level of probabilities on the values of another node upon knowing the value for a common child in the network's digraph. For example, a negative product synergy exhibited by nodes A and B with regard to the value c_0 for a common child C expresses that, given co, observing higher values for node A renders higher values for node B less likely, regardless of any other influences on B, and vice versa.

Definition 1.9 ***(Product synergies)*** *$X^{\delta}(\{A, B\}, C)$ holds in QPN G iff for a value c_0 of chance node C, X of C's prodecessors in G other than A and B, $a > \bar{a}$, and $b > \bar{b}$ of B :*

$$[F_c(c_0|abx) \cdot F_c(c_0|\bar{a}\bar{b}x)]\ R_{\delta}\ [F_c(c_0|a\bar{b}x) \cdot F_c(c_0|\bar{a}bx)].$$

where R_{δ} is $\leq, \geq$, or = as δ is +, -, or o, and $R_?$ is the complete relation (thus $S^?$ always holds).

The signs of the product synergies can be indicated over a dashed edge as it's shown in the figure below.

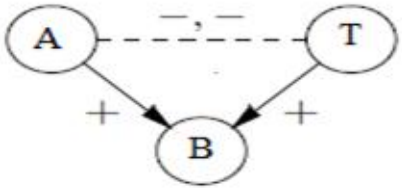


FIGURE 1.8 – *negative product synergy between A and T with regard to the value b for node B.*

Example

$$Pr(w|ul) \cdot Pr(w|\bar{u}\bar{l}) = 0.0081, \quad \text{and} \quad Pr(w|\bar{u}l) \cdot Pr(w|u\bar{l}) = 0.018.$$

$$Pr(w|ul) \cdot Pr(w|\bar{u}\bar{l}) \leq Pr(w|\bar{u}l) \cdot Pr(w|u\bar{l}) = 0.018.$$

We conclude that nodes U and L exhibit a negative product synergy with regard to the value w for node W.

1.3.4 The properties of symmetry, transitivity and composition

Qualitative relations of QPNs have the following properties, which allow for establishing influences between indirectly related nodes upon inference :

Symmetry : If B has an influence of sign δ $\{+, _, 0, ?\}$ on A, then A will have an influence with the same sign d on B.

Transitivity : If A has an influence of sign d1 on B, B has an influence of sign δ on C, and B does not have two incoming arrows on this trail, then this will result in an influence of A on C with a sign that is equal to the sign-product $\delta \oplus \delta$ (see Table 1.1) of the signs of the first and second influence.

Composition : The joint effect of influences along multiple incoming trails on a variable is defined by the operator $\otimes$ in the table 1.1.

1.3.5 Intercausal influence

Observations may induce intercausal effects when two a-priori independent variables A and B (two causes) are no longer independent given a specific value (evidence) for a common descendant (child). And this can be clearly seen on the product synergy.

Consider the following example : your brother is used to come back home at 20 :00 o'clock to watch the news but now it is very late and until 21 :00 he has not arrived. His late has two reasons : (A) he made an accident. (T) its late due to traffic. We assume that A and T are a-priori independent. However, given that you brother is too late observing in the TV news that there a dangerous accident near your neighborhood (an indicator for the cause A) makes it less likely that its late due to traffic. On the other hand, your neighbor came and he said that he spent an hour to get home because of traffic(an indicator for the cause T) makes it less likely that he made an accident. Therefore, variables that were a-priori

independent may become dependent given information for a common effect. The above equivalence then entails that the direction (sign) of the effect of actually observing a value for head-to-head node B on A, is the same as the direction of the effect of observing a value for A on B. This is related to the symmetry property.
We define the concept of intercausal influence more formally.

Definition 1.10 *Let $G = (V(G), A(G))$ be an acylic digraph and let Pr be a joint probability distribution on $V(G)$ such that G is an I-map for Pr. Le $A, B, C \in V(G)$ be nodes in G with $A \to C$, $B \to C \in A(G)$. Let $X = (\pi_G(B) \cup \pi_G(C)) - \{A\}$. Then, given the value c_0 for node C, node A exhibits a positive intercaucal influence on node B, written $Z_G^+(A, B, c_0)$, iff*

$$Pr(b|ac_0x) \geq Pr(b|\bar{a}c_0x)$$

for any combination of values x for the set of nodes X.

Negative, zero, and ambiguous intercausal influence again are defined analogously.

1.3.6 Reasoning with QPN

To reason with a QPN is to find the effect of new evidence on each node in terms of the sign of the change (increase or decrease) in networks and this is called inference in QPNs. We introduce a polynomial time algorithm for reasoning in QPNs, based on local sign propagation designed by (Druzdzel et Henrion 1993).

In essence, this sign-propagation algorithm computes the sign of influence along the active trails between the observed node and all other nodes, using the properties of symmetry, transitivity and composition. In this algorithm, messages are sent from a node to the so-called active neighbors of this node. The concept of active neighbors was adopted in order to implement d-separation at the local level. We give a slightly adapted version of the definitions presented by (van Kouwen et al. 2009).

Definition 1.11 *(Active neighbors). Let $G = (V(G), A(G))$ be an acyclic digraph. Let $A, B \in V(G)$ be nodes in G such that, during the inference, node A receives a message from node B. Let $O \subseteq V(G)$ be the set of observed nodes in G. Let $X = \{X_i | X_i \in \sigma(A)$ and $\sigma^*(X_i) \cap O \neq \emptyset\}$ be the set of children of A with an observed descendant.*

If $B \to A \in A(G)$ then the active neighbors of A consist of the set $(\sigma(A) \cup \pi(X)) - (\{A\} \cup O)$ otherwise, the active neighbors of A consist of the set $(\pi(A) \cup \sigma(A) \cup \pi(X)) - (\{A\} \cup \{B\} \cup O)$.

The sign-propagation algorithm takes as input a qualitative probabilistic network Q, a set Observed of previously observed nodes, a node O for which an observation has become available, and the sign of the current observation, that is, either a '+' for the value true or a '' for the value false. Prior to the propagation of the new observation, the node signs sign[V_i] for all nodes V_i are set to '0'. For the currently observed node O the appropriate sign is then entered into the network. The observed node

updates its node sign to the sign-sum of its original sign and the entered sign. It thereupon notifies all its active neighbors that its sign has changed, by passing to each of them a message containing a sign ; this sign is the sign-product of the node's current node sign and the sign *linksign* of the influence associated with the arc or edge it traverses. In each message it also records the origin of the message. A node to that receives a message, updates its node sign to the sign-sum of its current node sign sign[to] and the sign messagesign from the message it receives. The node then sends a copy of the message to all its active neighbors that need to update their node sign. In doing so, the node changes the sign in each copy to the appropriate sign and adds itself as the origin of the copy. The information about the origin of the copy is added to the information about the origin of the message from which the copy resulted ; as this process is repeated throughout the network, therefore, the trails along which messages have been passed are recorded.

We show the original algorithm as it was reformulated in pseudo-code.

procedure PropagateObservation$(Q, O, sign, Observed)$:
  for each $V_i \in V(G)$
  do sign$[V_i] \longleftarrow' 0'$;
  PropagateSign$(\varnothing, O, O, sign)$.

procedure PropagateSign$(trail, from, to, messagesing)$:
  sign$[to] \longleftarrow$ sign$[to] \oplus messagesing$;
  $trail \longleftarrow trail \cup \{to\}$;
  for each active neighbour V_i of to
  do $linksign \longleftarrow$ sign of (induced) influence between to and V_i;
    $messagesign \longleftarrow$ sign$[to] \otimes linksign$
    if $V_i \notin trail$**and** sign$[V_i] \neq$ sign$[V_i] \otimes messagesign$
    then PropagateSign$(trail, to, V_i, messagesing)$.

The sign-propagation algorithm for probabilistic inference in a qualitative network.

This sign-propagation algorithm is guaranteed to visit each node at most twice, since the node-sign of a node can change at most twice (once from '0' to '+' or '', then only to ' ?') ; the algorithm's computational complexity is linear in the number of arrows.
(van Kouwen et al. 2009) showed undesired results may occur with the original sign-propagation algorithm. The first and most serious type is that the algorithm calculates incorrect signs. The second type is that the algorithm calculates unnecessary ambiguity. They proposed so adapted sign-propagation algorithms where three adaptations have been made. First, they used an attribute "oppdir_sign" for nodes to capture the influences that entered this node through an outgoing arrow, in the opposite direction of the arrow. That second adaptation prevents irrelevant information from being passed through a head-to-head node's node-sign. The third prevents the premature halting of message-passing.

The adapted algorithm (van Kouwen et al. 2009) is as follow :

procedure PropagateObservation($QPN, O, sign, Observed$)
  for each $V_i \in V(G)$
  do sign$[V_i] \longleftarrow' 0'$; oppdir_sign$[V_i] \longleftarrow' 0'$
  PropagateSign($\varnothing, O, O, sign$)

procedure PropagateSign($trail, from, to, messagesing$)
  sign$[to] \longleftarrow$ sign$[to] \oplus messagesing$
  if $to \notin \sigma(from)$
  then oppdir_sign$[to] \longleftarrow$ oppdir_sign$[to] \oplus messagesing$
  $trail \longleftarrow trail \cup \{to\}$
  for each active neighbour V_i of to
  do $linksign \longleftarrow$ sign of (induced) influence between to and V_i
    if $to \in \sigma(from)$
  then $messagesing \longleftarrow$ oppdir_sign$[to] \otimes linksign$
  else $messagesing \longleftarrow$ sign$[to] \otimes linksign$
    if $V_i \notin trail$ **and not** SingsEqual($to, V_i, messagesign$)
    then PropagateSign($trail, to, V_i, messagesing$)
  function SingsEqual($to, V_i, messagesign$) : boolean
  $singsequal \longleftarrow$ **False**
  if $to \in \sigma(V_i)$
  then if oppdir_sign$[V_i] =$ oppdir_sign$[V_i] \oplus messagesing$
    then $singsequal \longleftarrow$ **True**
  else if sign$[V_i] =$ sign$[V_i] \oplus messagesing$
    then $singsequal \longleftarrow$ **True**
  return $singsequal$

Adapted sign-propagation algorithm

1.3.7 Extension of QPNs

(Renooij et al. 2003) defined a kappa-enhanced qualitative probabilistic network investigating the use of order-of-magnitude kappa values to capture strengths of influences in a qualitative network to provide a probabilistic reasoning in a qualitative way. The kappa calculus provides for a semi-qualitative approach for reasoning with uncertainty. In the kappa calculus, degrees of (un)certainty are expressed by a ranking κ. Kappa rankings can be interpreted as order-of-magnitude approximations of probabilities. Formally the order of magnitude as it's defined by (Renooij et al. 2003)

$$\kappa(x) = n \quad \text{iff} \quad \epsilon^{n+1} < Pr(x) < \epsilon^n.$$

Note that higher probabilities are associated with lower ?-values ; for example,

$$\kappa(x) = 0 \quad \text{if} \quad Pr(x) = 1, \quad \text{and} \quad \kappa(x) = \infty \quad \text{iff} \quad Pr(x) = 0.$$

[Renooij et al] defined the κ-value of an influence of A on B for a certain x :

$$\kappa(I_x(A,B)) = n \quad \text{iff} \quad \epsilon^{n+1} < |Pr(b|ax) - Pr(b|\bar{a}x)| < \epsilon^n.$$

We then define the strength factor associated with the influence of A on B to be an interval [p,q] such that

$$p \geq \max_x \kappa(I_x(A,B)) \quad \text{and} \quad 0 < q < \min_x \kappa(I_x(A,B)).$$

and each κ expresses an order of magnitude in terms of the same base. We associate strength factors with positive and negative influences ; zero and ambiguous influences are treated as in regular qualitative probabilistic networks.

In a kappa-enhanced qualitative probabilistic network we write $S^{\delta[p,q]}(A,B)$ to denote a qualitative influence of node A on node B with sign δ and strength factor $[p,q]$ in such a network. Their kappa-enhanced networks require only a number of kappa values that is linear in the number of parents for each node. Strength of influences in a kappa-enhanced network are expressed in terms of the base chosen for the network : the influence of node A on node B has strength factor [p,q] if for all x

$$\epsilon^{p+1} < |Pr(b|ax) - Pr(b|\bar{a}x)| < \epsilon^q.$$

1.4 Ontology

1.4.1 Introduction

Knowledge-based systems pose special requirements for interoperability. Such systems operate on and communicate using statements in a formal knowledge representation. They ask queries and give answers. They take "background knowledge" as an input. And as agents in a distributed AI environment, they negotiate and exchange knowledge. For such knowledge-level communication, we need conventions at three levels : representation language format, agent communication protocol, and specification of the content of shared knowledge. In this context ontology is created as a unifying framework to solve many problems in various domains (e.g., linguistics, semantic web, and bioinformatics). Ontology is the term used to refer to the shared understanding of some domain of interest. For the AI community, ontology is an explicit specification of a conceptualization (Gruber 1995). That is, ontology is a description of a set of representational primitives with which to model an abstract model of a knowledge domain.In particular there so are no standard definition for ontology. It is defined as the area of use.

1.4.2 Definition

Formally, ontology will be defined by :

Definition 1.12 *$C_p = \{cp_1, \ldots, cp_n\}$ is the set of n concepts (classes) such that each cp_i has a set of k properties (attributes) $P_i = \{p_1, \ldots, p_k\}$.*

- *R is the set of binary relations between elements of C_p which consists of two subsets :*

  - *H_R which describes the inheritance relations among concepts.*
  - *S_R which describes semantic relations between concepts.*
- *I is the set of instances, representing the knowledge base.*
- *A is the set of the axioms of the ontology. A consists of constraints on the domain of the ontology that involve C_p, R and I.(i.e. a set of constraints like must, must not, should, should not, etc).*

We use the same domain that we used to illustrate the example of the cognitive map and Figure 1.9 shows the clarifications that we made on the cognitive map shown in Figure 1.2 to achieve this ontology.

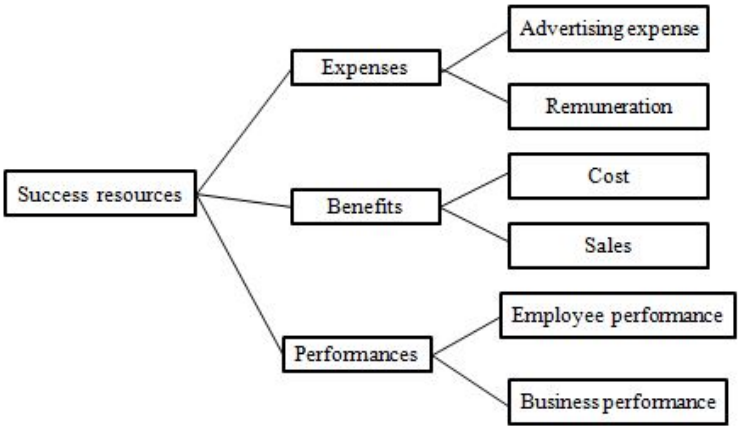


FIGURE 1.9 – *Example of ontology.*

1.4.3 Use of ontology

In this section we review and elaborate the motivations for ontology. In doing so we characterize the space of uses for ontology. (Uschold et Gruninger 1996) sub-divide the space of uses for ontologies into the following three Categories :

Communication **:** Recall that ontologies reduce considerably terminological confusion by providing a unifying framework within an organization. In this way, ontologies enable shared understanding and communication between people with different needs and viewpoints arising from their particular contexts.

Inter-operability **:** Many applications based on ontologies address the issues of interoperability in which different information management systems are deployed and different system-users need to exchange data using various software tools. The major contribution for the recourse to ontologies in domains such as enterprise modeling and multiagent architectures is the creation of a shared understanding of common domains allowing applications to agree on the terms that they are using when communicating. Hence, ontologies, if shared among the interoperating applications, allow a semantical data exchange between these tools.

Systems engineering specification **:** A shared and consistent understanding of problems and tasks at hand can assist in the specification of

software engineering project. In this way, software engineering ontologies are developed in order to represent and communicate over software engineering knowledge.

The development of such "software engineering domain ontology" will allow us to :

- share and reuse all knowledge accumulated until now in the Software Engineering field.
- open new avenues to automatic interpretation of this knowledge.

1.4.4 construction of an ontology

In this section we discus a skeletal methodology for building ontologies. Before starting building we have to identify the purpose and the scope of an ontology yet to be constructed ; It is important to be clear about why the ontology is being built and what its intended uses are. (Uschold et Gruninger 1996) defined three aspect to build an ontology :

***Capture* :** first we have to identify key concepts and relationships in the domain of interest. Then we have to produce precise unambiguous text definitions for such concepts and relationships and identify terms to refer to such concepts and relationships and finally agreeing on all of the above.

***Coding* :** by coding we mean explicit representation of the conceptualisation captured in the previous stage in some formal language. This will involve :

- committing to the basic terms that will be used to specify the ontology (e.g. class, entity, relation) ; this is often called a meta-ontology ;
- choosing a representation language (which is capable of supporting the meta-ontology) ;
- writing the code.

***Integrating existing ontologies* :** It is easy enough to identify synonyms and to extend an ontology where no concepts readily exist. However when there are obviously similar concepts defined in existing ontologies it is rarely clear how and whether such concepts can be adapted and reused. That's why many tools were developed to solve this problem.

1.5 Comparative study

This section represents a comparative study between four models. In fact in this table we show most important points of similarities and differences such as the structure, content, nature of nodes and links, etc. After this chapter we distinguish between purely probabilistic models (Bayesian networks) and purely qualitative models (Cognitive Maps) and we discovered that Qualitative Probabilistic Networks are a bridge between both. On the other hand we have adopted ontologies as a support model that helps to enrich whether qualitative networks or quantitative ones.

	Cognitive Map	**Bayes Nets**	**QPNs**	**Ontology**
Structure	Networks	Network	Network	Hierarchical structure starting from a general idea(concept)
Contents	personal understanding of the influences between concepts of a problem	personal understanding of the conditional independencies between concepts of a problem	personal understanding of the influences between concepts of a problem	personal understanding of semantic relationships between concepts
Characteristic Of nodes	Nodes contain concepts	Nodes contain random variable	Nodes contain ordered variable	Nodes contain classes (concepts)
Characteristic Of links	Links of influences	Links of conditional independencies. For causal Bayes nets : causal links	Links of influence	Semantic relationships between classes
Comprehension Difficulties	Understanding problem : a tendency to provide a personal vision.	Probabilities make understanding easier	Understanding problem : a tendency to provide a personal vision.	Understanding problem : a tendency to provide a personal vision
Construction Difficulties	Easy (few relations)	Complex because several representatives can be constructed	Capturing from an expert	Easy (few relations)

1.6 Existing bridges

1.6.1 Interaction between CM & Causal BNs

Cognitive maps are powerful graphical models for knowledge representation. They offer easy means to express individual's judgments, thinking or beliefs about a given problem. However, drawing inferences in cognitive maps, especially when the problem is complex, may not be an easy task. The main reason of this limitation in cognitive maps is that they do not model uncertainty with the variables. In this context is the work of (Nadkarni et Shenoy 2001) whose described a new graphical structure called Bayesian causal maps (BCM for short) to represent and analyze domain knowledge of experts. A Bayesian causal map is a causal map, i.e., a network-based representation of an expert's cognition. It is also a Bayesian network, i.e., a graphical representation of an expert's knowledge based on probability theory. To realize their goal authors used the initial CM only to construct the structure of the BCM. They defined the

parameters (local probabilities) of the BCM from experts knowledge. This means that domain experts are in a first step solicited for the construction of the CM and second in the definition of the conditional probabilities.By against (Sedki et Bonneau De Beaufort 2012) proposed an extension of the CM formalism by enriching the causal relation with additional information (for example, information about the influence of each causal concept when the premise variable increases and decreases). Then, they proposed to transform the constructed CM into BCM. Concerning the structure of the BCM, they followed the procedure described by (Nadkarni et Shenoy 2001). Concerning the parameters of the BCM, they proposed to take into account the causal values associated to the relations of the constructed CM. Namely, the elicitation step of the BCM will be less tedious since the structure and the parameters of the BCM are defined from the constructed CM. Their BCM was used to analyze the problem of shells fishing activity. The CM is constructed from judgments and beliefs of several actors (fishermen) and then they transformed it into a BCM following the proposed approach.

(Le Dorze et al. 2013) proposed a definition of the propagated influence when the information about the strength of an influence is interpreted as a probability value. To justify the definition of this operator, they proposed a method to encode a cognitive map into a causal Bayesian network

1.6.2 Interaction between CM & QPNs

(Van Kouwen et al. 2009) proposed a new approach for decision and planning support called Quasta tool as an abbreviation for the phrase 'qualitative start' based on QPNs. The aim is to explore the practical applicability of Quasta for participatory problem structuring through computer-supported cognitive mapping. The tool, allows qualitative exploration of scenarios and simultaneous forecasting and backcasting.

The Quasta tool :

- helps stakeholders to become aware of causal relationships ;
- helps in exploring possible scenarios ;
- identifies the need for further (quantitative) knowledge ;
- has a low threshold for nontechnicians.

QUASTA combines the problem structuring functionality of traditional CM with the analytical value of the QPN modeling technique. So instead of having several different interpretations from a simple reading of a cognitive map QUASTA define a unique direction of reading by changing directions of influences of variables automatically. To explain this phenomenon (**?**) proposed the example of scenario of traffic jams.

1.6.3 Interaction between CM & Ontology

The ontology helps the designer to find relevant concepts to build the map. In addition, it helps the user to understand concepts of the map. An ontological cognitive map is the association of ontology and a cognitive map. (Jung et al. 2003) and (Poignonec 2006) associated an ontology to a cognitive maps. They used an ontology as a tool for measuring similarity

between concepts in order to compare or merge a set of maps.

(Jung et al. 2003) defined ontology as a tree of concepts. (Jung et al. 2003) defined an ontological cognitive map as a map in which each concept is defined in the ontology. The objective of (Jung et al. 2003) was to determine the effect of a concept of a map on another concept of another map representing a different field. Jung proposed to connect different maps together by their common concepts. This operation is through the ontology concepts that are semantically equivalent when they are in different maps, and they used these concepts as bridges between cards.

(Chauvin et al. 2011) introduced the notion of ontological cognitive map that combines ontology and a map to organize cognitive concepts. To facilitate the understanding of a map, the user can get a simplified view of this card on a scale that he has chosen. A profile can be created to build views corresponding to the objectives of a specific user. If a card is handled by different users, their combined profiles used to build a shared view.

1.6.4 Interaction between BNs & QPNs

(Doan et al. 1999) presented a formalism called hybrid Bayesian networks that combines the traditional formalism with that of qualitative probabilistic networks. Their formalism allows a model builder to start by specifying only qualitative influences and then add quantitative information as it is available and as time permits. The modeling environment can infer bounds on unspecified probability values based on those specified and on the type of qualitative influence. They have built algorithms and implemented them in a prototype program to calculate the bounds of missing data in a single CPT of a child node with a parent node and its influence type.

(Lucas 2005) deployed both causal independence and QPNs in developing and analysing a collection of qualitative, causal interaction patterns, called QC patterns. These are endowed with a fixed qualitative semantics, and are intended to offer developers a high-level starting point when developing Bayesian networks.

1.6.5 Interaction between BNs & Ontology

(Ding et Peng 2004) developed a probabilistic framework for automatic ontology mapping by proposing an approach when they augmented the OWL language to allow additional probabilistic markups so that probability values can be attached to individual concepts and properties. In this framework, ontologies (or parts of them) are first translated into Bayesian networks, then the concept mapping is realized as evidential reasoning by developing a set of rules to translate a probability-annotated ontology into a BN structure.

(Laskey et da Costa 2005) presented Multi-entity Bayesian networks (MEBN), a formal system that integrates First Order Logic (FOL) with Bayesian probability theory. MEBN extends ordinary Bayesian networks to allow representation of graphical models with repeated sub-structures,

and can express a probability distribution over models of any consistent. Their approach called PR-OWL. Based on this approach (Carvalho 2011) have recently proposed a new syntax and semantics, defined as PR-OWL 2.0, which improves compatibility between PR-OWL 1.0 and OWL in two important respects. First, PR-OWL 2.0 formalizes the association between random variables from probabilistic theories with individuals, classes and properties from OWL. Second, PR-OWL 2.0 allows values of random variables to range over OWL datatypes.

(Yang et Calmet 2005) proposed an ontology-driven model, which integrates Bayesian Networks (BN) into the Ontology Web Language (OWL) to preserve the advantages of both. This model makes use of probability and dependency-annotated OWL to represent uncertain information in BN structures. These extensions enhance knowledge representation in OWL and enable agents to act under uncertainty. One of the main advantage of OntoBayes model is that it can preserve the ability to express meaningful knowledge in very large complex domains and extent ontologies to probability-annotated OWL to facilitate meaningful knowledge representation in uncertain systems.

All these works cited above are to build an ontology starting wit or based on a Bayesian network. There are other existing works where they constructed BN using ontologies where concepts were transformed into nodes, concepts attributes into CPTs and inheritance relation into links. In this context was the work of (Devitt et al. 2006) and (Fenz et al. 2009), but main difference between them is that (Fenz et al. 2009) constructed the BN directly from existing domain ontologies and do not require any BN-specific ontology extensions.
More recently, it was the work of (Ben Messaoud 2012) who investigated ontology for semantical causal discovery to construct a causal Bayesian network then based on this causal BN already constructed to discover new relations in the ontology by capturing some new causal relations to propose during the ontology evolution process.

Incremental design of Graphical knowledge-based systems

2

Sommaire

2.1 Description of the construction process 37
2.1.1 Phase I : Initial CM 37
2.1.2 Phase II : from CM to QPN 40
2.1.3 Phase III : from QPN to CBN 46
2.1.4 Phase IV : from CBN to QPN 48
2.2 process extension 48
2.2.1 General description of the principle 48
2.2.2 Hierarchy concept : helps in visualization 49
2.2.3 Semantic relations : help in construction 50

In this chapter we introduce a method to build knowledge based systems incrementally for reasoning and visualization. We use methods that exist and that we have already illustrated in the previous section enjoying their benefits and offering solutions to their limitations and drawbacks. We propose a global architecture that part from the construction of an initial cognitive map using an elicitation process includes several domain experts and arriving to causal Bayesian Network. We also propose methods to show how to proceed inversely, that's to say, how to transform a causal Bayesian Network to QPN for easier visualization. Then we propose some semantic exension to improve the visualization and the construction of our structures. The Figure 2.1 summarizes the approach by explaining the purpose of each phase ; if it is for reasoning or for visualization.

2.1 Description of the construction process

2.1.1 Phase I : Initial CM

Construction

The procedure consists firstly in constructing a cognitive map. There are two major steps in constructing cognitive maps. These steps are shown

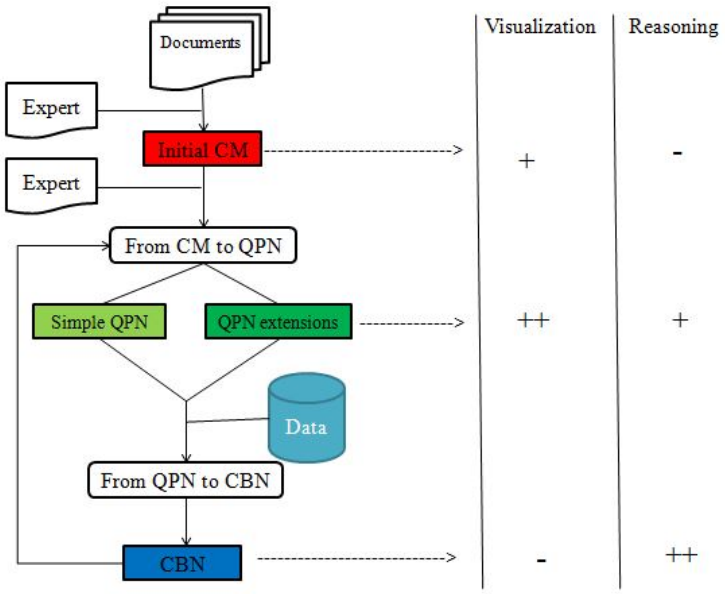


FIGURE 2.1 – *Schematic overview of the approach.*

and developed in our first chapter. First, information's source selection, this step consists to collect data from documents to define the different concepts. After selecting the domain that we will analyze and visualize an exploratory approach yields a richer understanding of the domain that individuals engage in decision making as well as helps us to get important insight into the general knowledge that individuals have on the domain being evaluated.

The experts elicitation step involves collecting causal relations those that explicitly contain a cause-effect relationship that links two different concepts through a causal connector such as 'if-then', 'because', 'so', 'as' etc. In this step, one way to develop a list of words to define causal connectors and to identify the causal relations and then recognizing these connectors in the relations is to use multiple experts.

Once the causal relations are identified, they are transformed into causal phrases to describe the influence between concepts by the language used by the expert (increase, decrease, no effect). Because the acquisition of signs require less effort and less time to spend with expert we propose therefore to start with a cognitive map that reflects the influences between concepts in the form of the (+) and of (-) then we try to enrich it step-by-step during our process. Once the list of causal links is defined, the coding scheme is derived from the cause and effect phrases.

Throughout this chapter we will illustrate the different stages of construction of our system by declining and enriching an example proposed by (Chauvin et al. 2011).

Identiying causal links

- High cloudy and rain can lead to an increase in the bad visibility.
- If there is rain then the road is so slippery.
- Because of rain there was a lot of fog.
- Thanks to the highway and the secondary road so there is no sinuous road.
- the number of secondary roads can be reduced by the construction of new highways.
- construction of new highways would negatively affect the silence in the nearby nature conservation area.
- Bad visibility, slippery road and sinuous road can cause accident.

Cause	*causal connector*	*Effect*
cloudy	$\xrightarrow{\text{lead to +}}$	bad visibility
rain	$\xrightarrow{\text{if-then +}}$	bad visibility
rain	$\xrightarrow{\text{if-then +}}$	slippery road
rain	$\xrightarrow{\text{because of +}}$	cloudy
highway	$\xrightarrow{\text{so -}}$	sinuous road
highway	$\xrightarrow{\text{reduced by -}}$	secondary road
highway	$\xrightarrow{\text{negatively affect -}}$	silence,nature conservation area
highway	$\xrightarrow{\text{reduced by -}}$	fine for drinving fast
secondary road	$\xrightarrow{\text{can cause +}}$	sinuous road
Bad visibility	$\xrightarrow{\text{can cause +}}$	accident
slippery road	$\xrightarrow{\text{can cause +}}$	accident
fine for drinving fast	$\xrightarrow{\text{increase +}}$	accident
sinuous road	$\xrightarrow{\text{can cause +}}$	accident

Use for reasoning

The cognitive map in Figure 2.2 captures an issue which is typical for the densely populated urban areas. It represents the various factors that may influence the risk of an accident occurring on a road. We consider a discussion concerning the construction of a new highway : supporters of the new highway argue that the project would be an essential measure for addressing traffic problems, so that reducing the number of accidents while other decision makers might argue that the new highway might disturb the silence in the nature conservation area. A third group might disagree with this reasoning. This group could suggest that sinuous roads would be reduced by construction of a new highway. Another stakeholder group could expand on this statement by explaining that sinuous roads increase the likelihood of accidents. Another group suggests that accidents can be reduced by increasing the fine for driving fast. Another stakeholder group might disagree with this reasoning, contending that personal

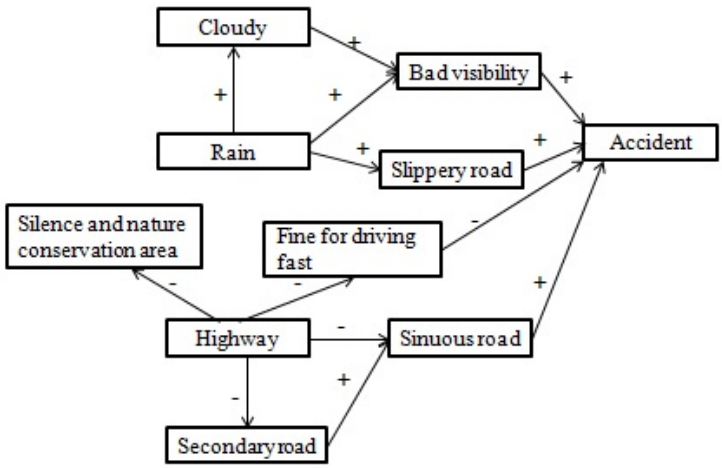


FIGURE 2.2 – *An example of the procedure for deriving an initial Cognitive Maps, result of phaseI.*

transportation is already very expensive, and therefore should not be burdened with additional costs. So that, CM provides qualitative reasoning. From simple reading we can identify several interpretations which make the analysis more difficult and several scenarios could be possible specially with complex CM where decision making become difficult and in making inferences become a nontrivial task because our initial CM do not model strength impact along the links and their variables have the same level of certainty. So the power to prescribe decisions can also be justified by an appeal to qualitative probabilistic networks.

2.1.2 Phase II : from CM to QPN

Construction

A qualitative approach for computer-supported analysis, such as QPN described in section 1.3 in which no strengths have to be specified allows us to use CM to make inferences for decision making. In this section, we sketch a procedure for constructing a simple QPN starting from the CM already constructed based on properties of QPN and CM. In this phase of our process, we conducted two follow-up interviews with expert. This interview was conducted to modify the initial cognitive map. That is to say, expert was requested to provide clarifications of the cause-effect relations by eliminating circular loop in the map based on the following instructions and proprieties that make the difference between the two graphs. The purpose of the second interview was to generate other extensions of the simple structure of QPN adding additional information to the original QPN based on the first interview.

Step 1 : Eliminating circular relations :In previous chapter, we have developed the formalism qualitative probabilistic networks, which share some significant features with cognitive maps. QPNs have been described in detail in section 1.3. Like cognitive maps, QPNs are signed directed acyclic

graphs, with nodes denoting concepts and signed edges denoting abstract causal relations. However, CM provides several types of configuration. Among these configurations, we were interested in the circular relationship. A circular relationship is a system where the relations are organized in a sort of buckle. Unfortunately, these circular relations in the CMs violate the acyclic graphical structure required in QPNs. It is therefore essential to eliminate circular relations to make our cognitive map compatible with QPN. To complete this stage, additionally interview has been conducted with expert. Causal loops can exist for two following reasons

- **Distinguishing between deductive and abductive reasoning :**To draw conclusions, make predictions, and construct explanations. Breif to make inference, there are two basic methods of reasoningthat we will be introducing here : deductive and abductive approaches. Deductive reasoning starts with the assertion of a general rule and proceeds from there to a guaranteed specific conclusion. Deductive reasoning moves from causes to effects, i.e., in the direction of causation. For example, consider our cognitive map rain is regarded as cause, and the slippery road as effects. When it rains reasons that the road is slippery, here we are reasoning deductively. Abduction is a method of reasoning in which one chooses the hypothesis that would, if true, best explain the relevant evidence. That is to say from the effect to the cause. For example, when we, after observing the road slippery and wet we conclude that it rains. Otherwise, abduction allows the precondition "rain"" to be abduced from the consequence "slippery road". Here we are reasoning abductively. In abductive reasoning, unlike in deductive reasoning, the premises do not guarantee the conclusion. Graphically speaking, the difference between deductive and abductive reasoning underlying causal statements and their effects that the direction of reasoning is in the direction of arrows, as well as in the opposite direction. This is clear in the graph below. Deductive and abductive reasoning also called respectively forward and backward reasoning QPNs allow forward and backward reasoning. This means that, when using the technique for exploration of scenarios, QPNs can be used for both forecasting and backcasting. Backcasting is backward reasoning from goals to action plans : beginning with the specification of certain objectives for the future and outlining what actions need to be taken to work towards those objectives (Van Kouwen et al. 2009). However, Causal statements involving abductive reasoning are misrepresented in a cognitive map by an arc from effect to cause. Such misrepresentation can also lead to redundant circular relations between variables in the causal map creating loops (Nadkarni et Shenoy 2001).
- **Distinguishing between direct and indirect relations :**During our first interview of deriving our initial CM we don't distinguish between direct and indirect relationships between concepts. May be, because experts draw unconsciously direct link between concepts even if they are influenced through another concept. However, This distinction is important to identify conditional independencies for our second structure. A lack of distinction between direct and indirect relationship can create circular structure as shown in the Figure

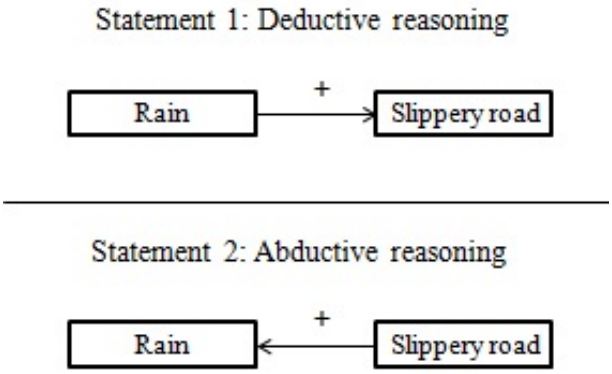


FIGURE 2.3 – *Distinguishing between deductive and abductive reasoning*

2.3. So, distinguishing between direct and indirect relationships is important namely in order to determine the conditional independencies between variables of the problem. Namely, circular form in a CM can be represented by a qualitative influence if the relation between variables is direct and if a given variable affects a second variable only through a third variable, then we can represent this by a dashed arrow reflecting a product synergy.

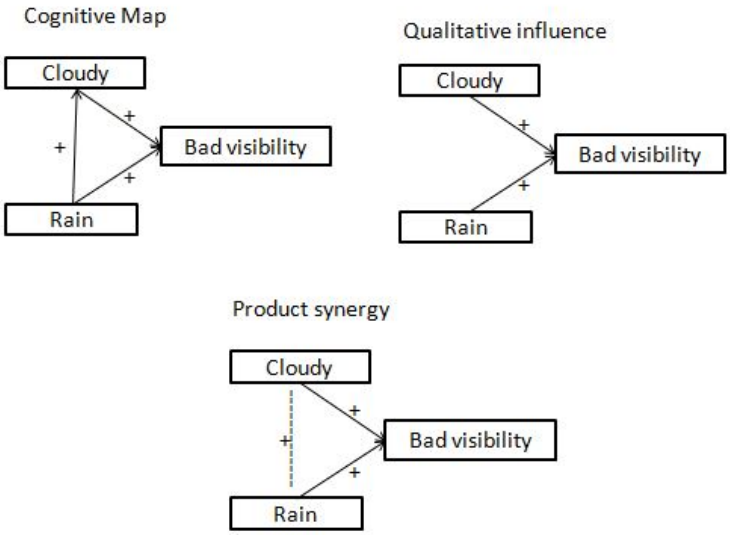


FIGURE 2.4 – *Distinguishing between direct and indirect relations*

In the Figure 2.4 we show two possible cases each case depends on the rethinking expert. If the relationship between rain and cloudy is redundant we can remove the links between these two variables creating a qualitative influence. If a given "rain" affects "cloudy" only through a value of "bad visibility", then we create a product synergy.
Figure 2.4 indicates a part of simple QPN with only qualitative influences.

Step 2 : Various extensions of QPN : Qualitative influences between variables, figure 2.5, are captured by simple signs without any indication of their strengths. As a consequence, any trade-off encountered during in-

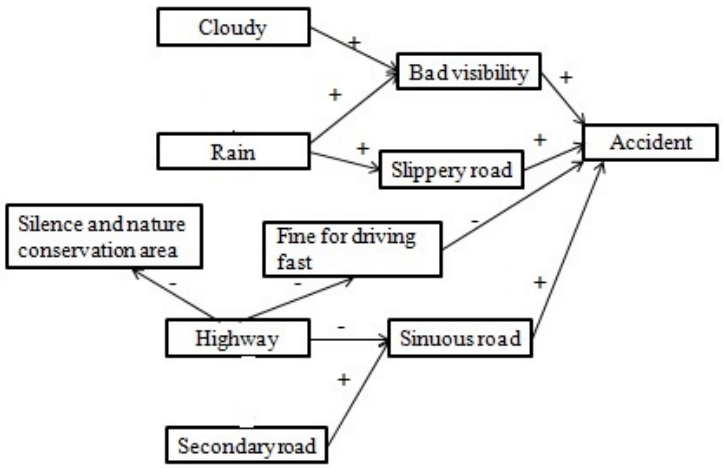


FIGURE 2.5 – *Simple QPN obtained from CM described in Figure 2.4*

ference will remain unresolved. We show how additional information can be collected from an expert to address the modeling issues discussed in step 1. In this step, we present other formalism for qualitative probabilistic networks that allows for a finer level of representation detail, which will enable the resolving of trade-offs to at least some extent. In this formalism, we enhance qualitative probabilistic networks by associating an indication of relative strength with their influences. We distinguish between strong and weak influences by adding additional information about the strength impact of a variable on other variable where numbers are specified or by identifying a probabilistic interval defining a semi-qualitative probabilistic network (Renooij et van der Gaag 2002)where Intervals have been used to indicate the uncertainty about or imprecision of the actual value of a probability.

To construct the figure 2.6 the subject determines the negative or positive effect of one concept on the others, with a fuzzy value for each interconnection, since it has been considered that there is a degree of causation between concepts. To have better results in the drawing of the enhanced QPN, we can use more than one expert. During the assigning of weights they must consider how strongly each concept X_i influences concept X_j. Each connection between concepts has a weight which ranges between [-1,1] and in this case it was determined arbitrarily.

To construct the figure 2.7 the subject determines intervals to indicate a range of differences in probability. We can also use Kappas as indicators of strength (Renooij et al. 2003) as an order-of-magnitude value to capture strengths of influences in a qualitative network already detailed in section 1.3. In this case strength factors provided by domain experts are intervals in [0,1] for positive influence, in [0,0] for zero influence and in -[1,0[for negative influence.

Other experts express the relationship between parents "Cloudy" and "Rain" as a positive influence of "Rain" to "Cloudy" observing a value of

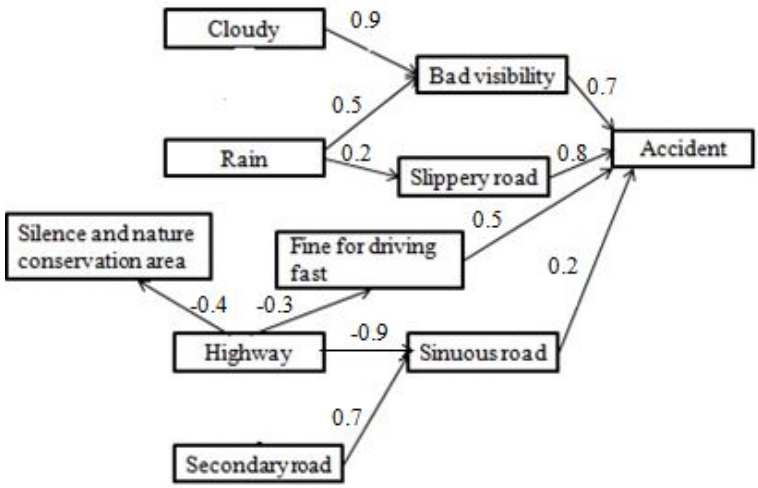


FIGURE 2.6 – *QPN with strength impact,possible improvment of QPN in Figure 2.5*

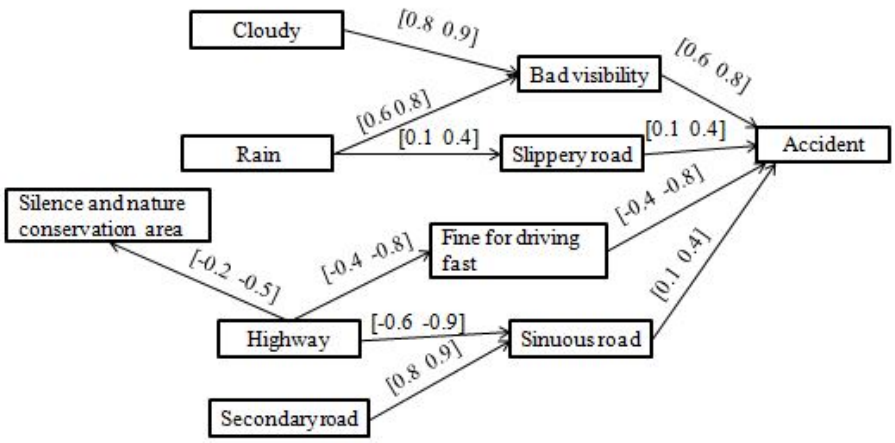


FIGURE 2.7 – *Interval QPN,possible improvment of QPN in Figure 2.5*

their common child "bad visibility". They explain that node child "bad visibility" has no parents other than "Cloudy"' and "Rain", by focusing on the situation of node "bad visibility" , nodes "Cloudy" and "Rain" then exhibit a product synergy. Consider C="Cloudy", R= "Rain", V="bad visibility" and V has values v for good visibility and $\bar{v}$ for bad visibility. Suppose that the nodes R and C quantified by the domain experts. They indicate that the conditional probabilities for node V are :

$$Pr(v|rc).Pr(v|\bar{r}\bar{c}) = 0.30, \quad \text{and} \quad Pr(v|\bar{r}c).Pr(v|r\bar{c}) = 0.28.$$
$$Pr(v|rc) + Pr(v|\bar{r}\bar{c}) \geq Pr(v|\bar{r}c) + Pr(v|r\bar{c}),$$

A similar observation holds with regard to the value $\bar{v}$ for node W. We substitute this probabilistic information in the qualitative network from Figure 2.8, thereby obtaining a positive product synergy $X^{+}(\{R, C\}, V)$. The same thing for parents H="highway" and S="secondary road" with

regard to the value of their child R="sinuous road" but in this case by creating a negative product synergy $X^{-}(\{H, S\}, R)$. We obtain Figure 2.8.

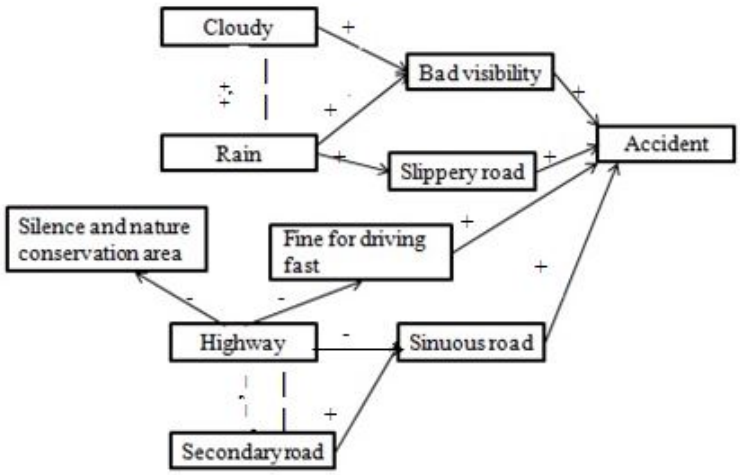


FIGURE 2.8 – *QPN with product synergies,possible improvment of QPN in Figure 2.5*

Use for reasoning

For probabilistic inference with a qualitative probabilistic network, a basic algorithm detailed in section 1.3.6. (van Kouwen et al. 2009) adopted the concept of QPN, and suggested to be used for computer-supported reasoning to resolve the problem of ambiguity and this also detailed in section 1.3.6. However, qualitative probabilistic networks, due to the high level of abstraction, do not provide for weighing influences with conflicting signs so that (Renooij et van der Gaag 2002) introduce interval-propagation algorithm that operates upon the associated interval network. Their algorithm is closely related to the based sign-propagation algorithm discussed in section 1.3.6 and is based on the idea of propagating intervals over arcs.

We consider the qualitative probabilistic network shown in Figure 2.8. Suppose that each node represents a binary variable. The qualitative influences between the nodes in the network are indicated over the digraph's arcs. We assume that the value of node "bad visibility" was observed to be true and that it has induced a positive intercausal influence between the nodes "rain" and "cloudy", indicated in the figure by the dashed edge between these nodes. In addition, there is a product synergy capturing the intercausal effects between "highway" and "secondary road" upon observing a common effect "sinuous road". Now suppose that we are interested in the effect of observing the value true for node "accident" upon all the other nodes in the network, in the presence of the previous observation for nodes "bad visibility" and "sinuous road". Suppose that positive evidence for node "accident" is entered into the network. Prior to the inference, the node signs for all nodes are set to 0. Node "accident" updates its node sign to the sign $0 \oplus + = +$. Then, the sign-propagation algorithm will start sending messages to the active neighbors of "accident" : "slippery road" and "fine for driving fast". No message will be sent to nodes "bad visibility" and "sinuous road" as they are observed. For nodes "slippery road"

and "fine for driving fast", node "accident" computes the sign +⊗+ = +. It thereupon sends two messages containing the computed '+' to nodes "slippery road" and "fine for driving fast". Now, nodes "slippery road" and "fine for driving fast" update their node sign to 0⊕+ = +. "slippery road" subsequently computes the sign +⊗+=+ to be sent to node "rain". After updating its node sign to 0⊕+=+, node "rain" sends a message with the sign +⊗+ = + to its active neighbors "cloudy" which updates its node sign to 0⊕+=+. Node "rain" does not send any message to node "bad visibility" as this node has been observed. The induced neighbour "fine for driving fast" of the node "accident", send a message sign +⊗-=- to "highway". After this, node "highway" will have the sign -+0=-. In addition, node "highway" sends a message with the sign -⊗-=+ to "secondary road" and "silence and nature conservation area". No message will be sent to nodes "sinuous road" it's observed. "Secondary road" and "silence and nature conservation area" update their node signs to 0⊕+=+. The process of message passing now halts.

2.1.3 Phase III : from QPN to CBN

Construction

To define the parameters of a Bayesian network in the case of complete data we obviously have two methods. The first method is the simplest and most widely used is the statistical estimation that is to estimate the probability of an event by the frequency of occurrence of the event in the database. In this case we do not use a domain expert. This approach, called maximum likelihood :

$$\hat{P}(X_i = x_k | pa(X_i) = x_j) = \hat{\theta}MV_{i,j,k} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \tag{2.1}$$

With $N_{i,j,k}$ is the number of events in the database for which the variable X_i is in the state x_k and his parents are in the configuration x_j.

In the second method we use the principle of Bayesian estimation is somewhat different. It is to find the most likely parameters θ as the data were observed using a priori on parameters. This approach is called maximum a posteriori (MAP) :

$$\hat{P}(X_i = x_k | pa(X_i) = x_j) = \hat{\theta}^{MV}_{i,j,k} = \frac{N_{i,j,k} + \alpha_{i,j,k} - 1}{\sum_k (N_{i,j,k} + \alpha_{i,j,k} - 1)} \tag{2.2}$$

With $\alpha_{i,j,k}$ are the parameters of the Dirichlet distribution associated with to prior law $\hat{P}(X_i = x_k | pa(X_i) = x_j)$.
In this section, we sketch a procedure for constructing Bayesian Network. The approach consists of two major steps. In the first step, we propose to build our network from the transformation of our QPN that is already constructed. We use the information provided by the expert to build our QPN (such as influences, their types, their importance degrees, etc.) and we translate them to the CPTs. In this case we already have several other sub-cases depending on the type of QPN that we will transform. The first

sub-case that we can transform the simple QPN with only '+' and '-' influences. This procedure is similar to existing algorithms for computing bounds or intervals of the desired probability distributions or even the exact distributions of Bayesian networks that we already mentioned in section 1.6.4. But these methods either they propose a definition of the propagated influence when the information about the strength of an influence is interpreted as a probability value, or they are based on calculating the bounds of missing data in a single CPT of a child node with a parent node and its influence type. They don't take into account the case of synergies of any subset of parent nodes. To illustrate the phenomenon, we consider the figure 2.8, just to give an idea of the result that we have to achieve after transformation of QPN to CBN.
For reasons of simplicity we will be limited by the illustration of CPT of just a part of the network.We suppose that the variable Rain='R' has two states : small and heavy, Cloudy='C' variable has two states : average and very, and Bad Visibility='BV' has two states :yes and no.

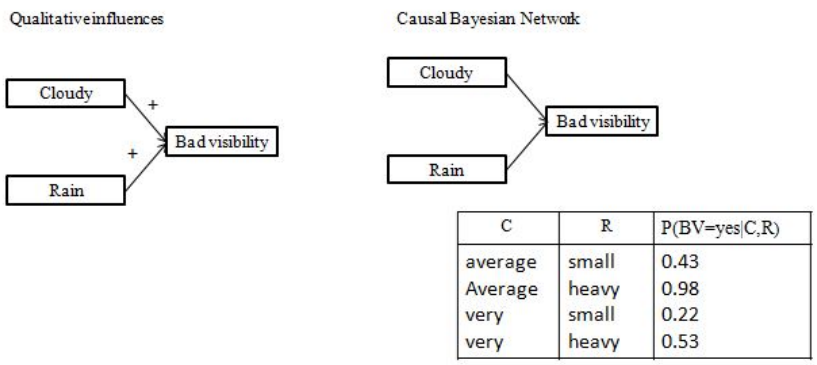


FIGURE 2.9 – *Example of result of the transformation of qualitative influences to table of probabilities*

The second sub-case when we have an enhanced QPN that is to say a QPN that is enriched by additional information such as : strength impact (+/-n), intervals probabilities, product and additive synergies. Researches are carrying out research for the case of synergies of any subset of parent nodes because this task is rather complicated.
In the second step we assume that we have a database that stores data about our network. We propose to use these data with information from our QPN already built giving rise to a Bayesian approach by considering our α_i are tables of probabilities that we already have obtained in the first step that is to say, from the transformation of the QPN to CBN. In both steps we end up with two different networks. The CBN of the second step is richest and most accurate and allows more efficient reasoning.

Use for reasoning

Once the structure and parameters of our Bayesian network are well defined we can now use this structure for inference and make decisions to solve a given problem basing on methods and reasoning algorithms that we have already discussed with more detail in section 1.2.3.

2.1.4 Phase IV : from CBN to QPN

Construction

A general CBN may contain a large number of variables with very complicated connections and many complicated CPTs. It is so difficult to be visualized directly especially for non-professional person. In fact the difficulty of interpreting probabilities weakens the ability to visualize and reflect on the domain represented by the CBN. Our proposal is to build from a causal BN a QPN that user can see and exploit easily. In relation to our process is to proceed inversely. For viewing the CBN which is built in the first step we can directly use our result QPN of Phase II. However to visualize the CBN built in the second stage, we need to turn it to a QPN because it is built not only from the information provided by our QPN but also from the frequencies of occurrence of events in the database.
In this case there are two suggestions, either by acting on the arrows that connect variables our networks. The intensity of the arrow highlights the importance of influence between each two nodes so we can have fine arrows to indicate less important influences or fatty arrows to indicate greater influences and this is already implemented with the Genie-Smile tool. But this method just highlights the importance of dependency between variables but we can't know if it is positive or negative influence. Another suggestion is to turn squarely the tables of probabilities to qualitative influences '+' and '-', product and additive synergies. This task is rather complicated since there is no program allows to such transformation, and this is another line of work that we can also study on proposing methodologies and processing among other tracks that we have identified throughout our process description.

Use for visualization

The basic belief that 'seeing' is a good way of understanding and generating knowledge. Humans have a very well developed sense of sight. So that, QPN is a powerful model can greatly promote the easy comprehension of the model because it represent a person's understanding of a topic by mapping concepts and their relationships. That's why in order to visualize our CBN structure more easily we propose to use the QPN obtained in second step of phase IV.

2.2 PROCESS EXTENSION

2.2.1 General description of the principle

We have previously shown that in order to learn CBNs, we chose to start with a simple structure to a more complex one by adding more out information and making changes to the structure of our network so that it will be adapted to each case that we want to study basing on causal statements, their types, their degrees of importance, etc. Therefore, every additional knowledge can improve the structure learning and therefore our model representation. In this section we propose to add another type of information wich is semantic information by integrating ontology

model. Ontology has two types of knowledge :the hierarchy of concepts called taxonomy and semantic relations between concepts. We propose to use these knowledge to improve respectively the construction and the visualization of our structures.

2.2.2 Hierarchy concept : helps in visualization

We can use the ontology to better visualize our various structures : CM, QPN and CBN. There are works that offered solutions to visualize each structure by using an ontological model. these methods are the interaction between ontology and each type of network and they are well detailed in section 1.6.3 and 1.6.5. In this case the ontology is used as a scale for providing a synthetical view of a map. The map is then easier to understand. Figure 2.9 displays the domain ontology that we can use to construct our ontological cognitive map.

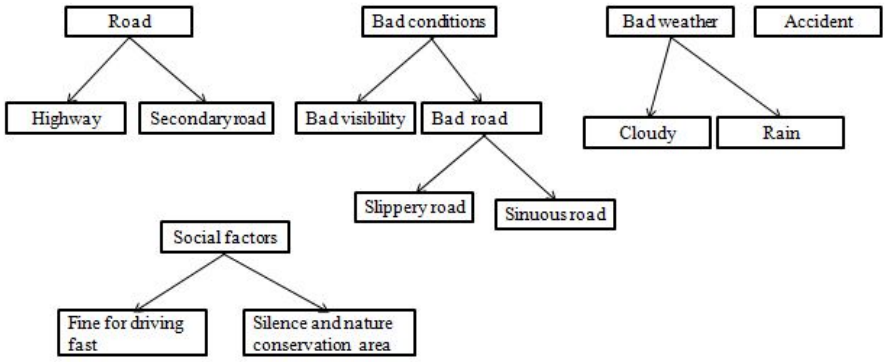


FIGURE 2.10 – *The Ontology.*

The integration of ontology guarantees an easier interpretation of map. Once our OCC is defined we can enrich it by adding strength impact or intervals of probabilities to obtain the various extensions of QPNs.

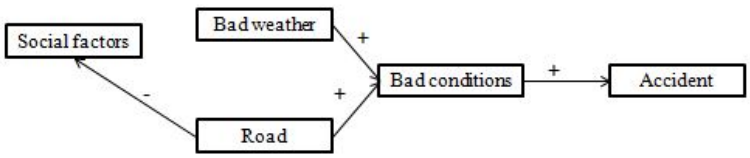


FIGURE 2.11 – *Ontological Conitive Map.*

Inserting ontology provided a simple representation to apprehend by all and providing different visualizations of knowledge. It promotes the participation of all stakeholders -expert or non-expert- in decision-making. However, it does not have very sophisticated reasoning mechanisms because to construct the OCC we have extracted potential relationships between the existing concepts without presenting the details of links.
We can also use the taxsonomy to better visualize our QPN. Our current focus is on linkage reducing and therefore reducing the number of questions asked to the expert. This is more important in case where we have a product synergy where the same parents are connected to more than two

children simultaneously,Figure 2.11, in this case to know the type of influence between parents observing a value of each child, it takes long time and requires several questions to be asked to the domain expert. Another undesired results may occur with this case is that we can obtain incorrect signs or ambiguous signs that is difficult to transform to probabilities even impossible. To overcome these pitfalls we use so ontology model. Namely, we can actively work on each of these components.

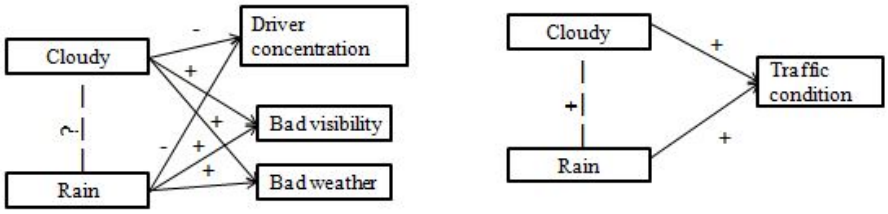


FIGURE 2.12 – *Result of insering ontology model in a product synergy.*

2.2.3 Semantic relations : help in construction

in this section we propose to use semantic relations of an ontology to build our structures. we assume that we have a knowledge of type "affect" or "affect positively/affect negatively", so how we can use these to build respectively our CM was our QPN ?.After careful study of the ontology domain, we also assume that we have discovered that we can extract knowledge about the hierarchy of influences such as "strong impact", "medium impact" and "low impact". This knowledge reflects an order of influences so it can help to build the QPN. Now, we assume that we have probabilistic ontologies, so the question arises how we can use it to build our CBN ?. All these questions can be future research paths. There are works that already used the probabilistic ontologies to build their CBN such as SemCaDo work that we talked about in our state of the art in the section 1.6.5.

Estimation of probabilities 3

Sommaire

3.1 Log-linear model approach 51
3.1.1 Links between $\mu - terms$ and CPTs 52
3.1.2 Links between $\mu - terms$ and semantic of QPN 57

In this chapter we focus on the Phase III of our process described in the previous section, especially on the deriving of causal Bayesian network parameters from causal influences provided by qualitative probabilistic network. Our goal in this section is to find the relationship between the semantics of the QPN and μ-terms of log-linear model by using examples of fragments of a QPN and tables of probabilities provided by an expert. For reasons of simplicity we begin with QPN with just positive or negative influences between binary variables. We start with an example of network composed of two nodes. Then, we try with graphs composed of two and three parents to highlight synergies provided by QPN.

3.1 Log-linear model approach

Log-linear models are the subject of several publications we cite for example (Bishop et al. 1975) and (Christensen 1997). A log-linear model is a mathematical model that takes the form of a function whose logarithm is a first-degree polynomial function of the parameters of the model, which makes it possible to apply (possibly multivariate) linear regression. These models are a far-reaching extension of Bernoulli models. Log-linear model is a technique used in statistics to examine the relationship between more than two qualitative variables with several modalities. The goal of log-linear analysis is to determine which model components are necessary to retain in order to best account for the data. Model components are the number of main effects and interactions in the model. There are two kind of log-linear model : simplest model and saturated model. The simplest model is the model where all the expected frequencies are equal. This is true when the variables are not related. The saturated model is the model that includes all the model components. This model will always explain the data the best, but it is the least parsimonious as everything is included. Each log-linear model can be represented as a log-linear equation.

For example, with three variables (A,B,C) the saturated model has the following log-linear equation :

$$\begin{aligned}\log(p(a,b,c)) =& \mu_0 + \mu_A(a) + \mu_B(b) + \mu_C(c) \\ &+ \mu_{AB}(a,b) + \mu_{AC}(a,c) + \mu_{BC}(b,c) \\ &+ \mu_{ABC}(a,b,c)\end{aligned} \tag{3.1}$$

Where μ , are terms of interactions, called μ-terms are functions of modalities of variables. Interactions where we have one varible correspond to main effects. Interactions where we have two varibles correspond to first order interactions, interactions where we have three varibles correspond to second order interactions.
We can have a simplest model implies conditional independence. We consider the model presented in the equation 3.1. We assume that B and C are independent this implies that $\mu_{BC}(b,c) = 0$.
A simplest log-linear model for this example could be :

$$\begin{aligned}\log(p(a,b,c)) =& \mu_0 + \mu_A(a) + \mu_B(b) + \mu_C(c) \\ &+ \mu_{AB}(a,b) + \mu_{AC}(a,c) \\ &+ \mu_{ABC}(a,b,c)\end{aligned} \tag{3.2}$$

If A,B and C are binary variables ; they have as value :0,1. The model can be written as follow :

$$\log(p(a,b,c)) = \mu_0 + \mu_A a + \mu_B b + \mu_C c + \mu_{AB} ab + \mu_{AC} ac + \mu_{ABC} abc \tag{3.3}$$

We note that the log-linear models support conditional independence so they can properly present a Bayesian network. (Corset 2003) defined a graphical log-linear model as follow :

Definition 3.1 *A log-linear model can be presented by a Bayesian network if, when the model contains all interaction terms of order 2 generated by an interaction term more larger, then the model contains the interaction of higher order.*

3.1.1 Links between $\mu - terms$ and CPTs

Example 1 : we consider the example in the Figure 3.1.
For this simple example with two binary variables the log-probability can

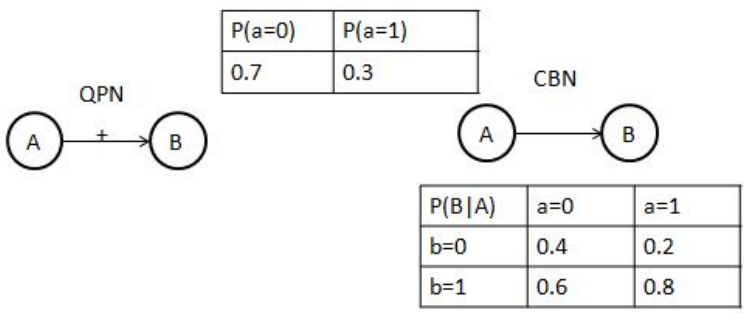


FIGURE 3.1 – *Example 1 of log-linear model.*

be written :

$$\log(p(a,b,c)) = \mu_0 + \mu_A a + \mu_B b + \mu_{AB} ab \tag{3.4}$$

In the case of binary variables the joint probabilities can be written as :

$$\begin{cases} \log P(0,0) &= \mu_0 \\ \log P(0,1) &= \mu_0 + \mu_B \\ \log P(1,0) &= \mu_0 + \mu_A \\ \log P(1,1) &= \mu_0 + \mu_A + \mu_B + \mu_{AB} \end{cases}$$

This is a system with four equations and four unknowns :μ_0,μ_A,μ_B and μ_{AB}. We can thus determine μ_B and μ_{AB} from the following conditional probabilities.

$$\begin{cases} \log P(B=0|A=0) &= \frac{1}{1+e^{\mu_B}} \\ \log P(B=0|A=1) &= \frac{1}{1+e^{\mu_B+\mu_{AB}}} \end{cases}$$

We assume that these two probabilities are provided by an expert, we can determine μ_B from the first conditional probability and we deduce μ_{AB} from the second conditional probability.

$$\mu_B = \log \frac{1}{P(B=0|A=0)} - 1$$

$$\mu_{AB} = (\log \frac{1}{P(B=0|A=1)} - 1) - \log(\frac{1}{P(B=0|A=0)} - 1)$$

However, to determine μ_0 et μ_A we need to provide more information such as information about P(A). We suppose that we have also P(A=0) then $P(A=0) = e^{\mu_0(1+e^{\mu_B})}$ from this equation we can determine $\mu_0 = \log \frac{P(A=0)}{1+\mu_B}$. To determine μ_A we calculate

$$P(A=0) + P(A=1) = 1 + e^{\mu_B} + e^{\mu_A}(1 + e^{\mu_B+\mu_{AB}}) = 1$$

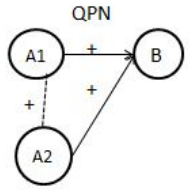
then

$$e^{\mu_A} = \frac{e^{-\mu_0} - 1 - e^{\mu_B}}{1 + e^{\mu_B+\mu_{AB}}}$$

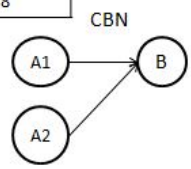
From this analysis we conclude that to determine μ_B and μ_{AB} we need to provide the table of probability $P(B|A)$ but to calculate μ_0 and μ_A we must have an information about P(A).

Example 2 : we consider now a simple network with two parents, Figure 3.2.

$$\begin{aligned} \log(p(a_1,a_2,b)) =& \mu_0 + \mu_A a_1 + \mu_A a_2 + \mu_B b \\ &+ \mu_{A_1B} a_1 b + \mu_{A_2B} a_2 b + \mu_{A_1A_2} a_1 a_2 \\ &+ \mu_{A_1A_2B} a_1 a_2 b \end{aligned} \tag{3.5}$$



P(A2|A1)	a1=0	a1=1
a2=0	0.4	0.2
a2=1	0.6	0.8



P(B|A1,A2)	a2=0	a2=0	a2=1	a2=1
	a1=0	a1=1	a1=0	a1=1
b=0	0.4	0.2	0.3	0.15
b=1	0.6	0.8	0.7	0.85

FIGURE 3.2 – *Example 2 of* log-*linear model with two parents.*

$$\begin{cases} \log P(0,0,0) &= \mu_0 \\ \log P(0,0,1) &= \mu_0 + \mu_B \\ \log P(0,1,0) &= \mu_0 + \mu_{A_2} \\ \log P(0,1,1) &= \mu_0 + \mu_{A_2} + \mu_B + \mu_{A_2B} \\ \log P(1,0,0) &= \mu_0 + \mu_{A_1} \\ \log P(1,0,1) &= \mu_0 + \mu_{A_1} + \mu_B + \mu_{A_1B} \\ \log P(1,1,0) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_{A_1A_2} \\ \log P(1,1,1) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_B + \mu_{A_1B} + \mu_{A_2B} + \mu_{A_1A_2} + \mu_{A_1A_2B} \end{cases}$$

$$\begin{cases} \log P(B=0|A_1=0,A_2=0) &= \frac{1}{1+e^{\mu_B}} \\ \log P(B=0|A_1=0,A_2=1) &= \frac{1}{1+e^{\mu_B+\mu_{A_2B}}} \\ \log P(B=0|A_1=1,A_2=0) &= \frac{1}{1+e^{\mu_B+\mu_{A_1B}}} \\ \log P(B=0|A_1=1,A_2=1) &= \frac{1}{1+e^{\mu_B+\mu_{A_1B}+\mu_{A_2B}+\mu_{A_1A_2B}}} \end{cases}$$

From these conditional probabilities we deduce respectively : μ_B, μ_{A_2B}, μ_{A_1B}, and $\mu_{A_1A_2B}$.
From the first and second example we conclude that a node with two parents requires more values to be determined.

$$\begin{cases} P(0,0) &= P(0,0,0) + P(0,0,1) &= e^{\mu_0}(1+e^{\mu_B}) \\ P(0,1) &= P(0,1,0) + P(0,1,1) &= e^{\mu_0+\mu_{A_2}}(1+e^{\mu_B}) + \mu_{A_2B} \\ P(1,0) &= P(1,0,0) + P(1,0,1) &= e^{\mu_0+\mu_{A_1}}(1+e^{\mu_B}) + \mu_{A_1B} \\ P(1,1) &= P(1,1,0) + P(1,1,1) &= e^{\mu_0+\mu_{A_1}+\mu_{A_2}+\mu_{A_1A_2}} \\ & & \times(1+e^{\mu_B+\mu_{A_1B}+\mu_{A_2B}+\mu_{A_1A_2B}}) \end{cases}$$

We suppose now that we have the joint probability $P(A2|A1)$ then, from these joint probabilities, we deduce respectively μ_0, μ_{A_2}, μ_{A_1}, and $\mu_{A_1A_2}$.
Example 3 : We consider now a simple network with three parents. the saturated log-linear model corresponds to the network showed in Figure 3.3 is :

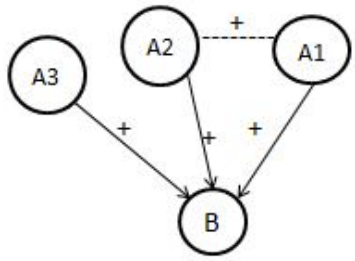


FIGURE 3.3 – *Example 3 of log-linear model with three parents.*

$$\begin{aligned}\log(p(a_1,a_2,a_3,b)) =& \mu_0 + \mu_A a_1 + \mu_A a_2 + \mu_A a_3 + \mu_B b \\ &+ \mu_{A_1B} a_1 b + \mu_{A_2B} a_2 b + \mu_{A_3B} a_3 b \\ &+ \mu_{A_1A_2} a_1 a_2 + \mu_{A_1A_3} a_1 a_3 + \mu_{A_2A_3} a_2 a_3 \\ &+ \mu_{A_1A_2B} a_1 a_2 b + \mu_{A_1A_3B} a_1 a_3 b \\ &+ \mu_{A_2A_3B} a_2 a_3 b + \mu_{A_1A_2A_3} a_1 a_2 a_3 \\ &+ \mu_{A_1A_2A_3B} a_1 a_2 a_3 b\end{aligned} \tag{3.6}$$

At this writing there are unnecessary interaction terms. According to the definition of synergy in a QPN product synergy involves only three variables then only terms of order 3 are useful. Terms $\mu_{A_1A_2B}$, $\mu_{A_1A_3B} a_1 a_3 b$, and $\mu_{A_2A_3B} a_2 a_3 b$ present the influence between each three variables. This information is presented in a QPN by a product synergy. That is to say, if we have a product synergy this term is different from zero and if there is no product synergy the term vanishes. If we return to our QPN in Figure 21 we have only one product synergy $\mu_{A_1A_2B}$. For synergies $\mu_{A_1A_3B} a_1 a_3 b = \mu_{A_2A_3B} a_2 a_3 b = 0$.
By eliminating unnecessary μ-terms, this gives :

$$\begin{aligned}\log(p(a_1,a_2,a_3,b)) =& \mu_0 + \mu_A a_1 + \mu_A a_2 + \mu_A a_3 + \mu_B b \\ &+ \mu_{A_1B} a_1 b + \mu_{A_2B} a_2 b + \mu_{A_3B} a_3 b \\ &+ \mu_{A_1A_2} a_1 a_2 + \mu_{A_1A_3} a_1 a_3 + \mu_{A_2A_3} a_2 a_3 \\ &+ \mu_{A_1A_2B} a_1 a_2 b + \mu_{A_1A_2A_3} a_1 a_2 a_3\end{aligned} \tag{3.7}$$

$$\begin{cases}
\log P(0,0,0,0) &= \mu_0 \\
\log P(0,0,0,1) &= \mu_0 + \mu_B \\
\log P(0,0,1,0) &= \mu_0 + \mu_{A_3} \\
\log P(0,0,1,1) &= \mu_0 + \mu_{A_3} + \mu_B + \mu_{A_3B} \\
\log P(0,1,0,0) &= \mu_0 + \mu_{A_2} \\
\log P(0,1,0,1) &= \mu_0 + \mu_{A_2} + \mu_B + \mu_{A_2B} \\
\log P(0,1,1,0) &= \mu_0 + \mu_{A_2} + \mu_{A_3} + \mu_{A_2A_3} \\
\log P(1,0,0,0) &= \mu_0 + \mu_{A_1} \\
\log P(1,0,0,1) &= \mu_0 + \mu_{A_1} + \mu_B + \mu_{A_1B} \\
\log P(1,0,1,0) &= \mu_0 + \mu_{A_1} + \mu_{A_3} + \mu_{A_1A_3} \\
\log P(1,0,1,1) &= \mu_0 + \mu_{A_1} + \mu_{A_3} + \mu_B + \mu_{A_1B} + \mu_{A_3B} + \mu_{A_1A_3} \\
\log P(1,1,0,0) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_{A_1A_2} \\
\log P(1,1,0,1) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_{A_1B} + \mu_{A_2B} + \mu_B \\
& \quad + \mu_{A_1A_2} + \mu_{A_1A_2B} \\
\log P(1,1,1,0) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_{A_3} + \mu_{A_1A_2} + \mu_{A_1A_3} \\
& \quad + \mu_{A_2A_3} + \mu_{A_1A_2A_3} \\
\log P(1,1,1,0) &= \mu_0 + \mu_{A_1} + \mu_{A_2} + \mu_{A_3} + \mu_B + \mu_{A_1B} + \mu_{A_2B} + \mu_{A_3B} \\
& \quad + \mu_{A_1A_2} + \mu_{A_1A_3} + \mu_{A_2A_3} + \mu_{A_1A_2B} + \mu_{A_1A_2A_3}
\end{cases}$$

$$\begin{cases}
P(B=0|A_1=0,A_2=0,A_3=0) &= \frac{1}{1+e^{\mu_B}} \\
P(B=0|A_1=0,A_2=1,A_3=0) &= \frac{1}{1+e^{\mu_B+\mu_{A_2B}}} \\
P(B=0|A_1=1,A_2=0,A_3=0) &= \frac{1}{1+e^{\mu_B+\mu_{A_1B}}} \\
P(B=0|A_1=1,A_2=0,A_3=1) &= \frac{1}{1+e^{\mu_B+\mu_{A_3B}}} \\
P(B=0|A_1=1,A_2=1,A_3=1) &= \frac{1}{1+e^{\mu_B+\mu_{A_1B}+\mu_{A_2B}+\mu_{A_1A_2B}}}
\end{cases}$$

From these conditional probabilities we can calculate μ_B, μ_{A_1B}, μ_{A_2B}, μ_{A_3B}, μ_{A_1B}, μ_{A_2B}, and $\mu_{A_1A_2B}$. For the remain unknowns we must calculate the following joint propabilities :

$$\begin{cases}
P(0,0,0) &= P(0,0,0,0) + P(0,0,0,1) \\
&= e^{\mu_0}(1+e^{\mu_B}) \\
P(0,1,0) &= P(0,1,0,0) + P(0,1,0,1) \\
&= e^{\mu_0+\mu_{A_2}}(1+e^{\mu_B}) + \mu_{A_2B} \\
P(1,0,0) &= P(1,0,0,0) + P(1,0,0,1) \\
&= e^{\mu_0+\mu_{A_1}}(1+e^{\mu_B}) + \mu_{A_1B} \\
P(0,0,1) &= P(0,0,1,0) + P(0,0,1,1) \\
&= e^{\mu_0+\mu_{A_1}}(1+e^{\mu_B}) + \mu_{A_1B} \\
P(1,1,0) &= P(1,1,0,0) + P(1,1,0,1) \\
&= e^{\mu_0+\mu_{A_1}+\mu_{A_2}+\mu_{A_1A_2}}(1+e^{\mu_B+\mu_{A_1B}+\mu_{A_2B}+\mu_{A_1A_2B}}) \\
P(1,0,1) &= P(1,0,1,0) + P(1,0,1,1) \\
&= e^{\mu_0+\mu_{A_1}+\mu_{A_3}+\mu_{A_1A_3}}(1+e^{\mu_B+\mu_{A_1B}+\mu_{A_3B}}) \\
P(0,1,1) &= P(0,1,1,0) + P(0,1,1,1) \\
&= e^{\mu_0+\mu_{A_2}+\mu_{A_3}+\mu_{A_2A_3}}(1+e^{\mu_B+\mu_{A_2B}+\mu_{A_3B}}) \\
P(1,1,1) &= P(1,1,1,0) + P(1,1,1,1) \\
&= e^{\mu_0+\mu_{A_1}+\mu_{A_2}+\mu_{A_3}+\mu_{A_1A_2}+\mu_{A_1A_3}+\mu_{A_2A_3}+\mu_{A_1A_2A_3}} \\
& \quad \times(1+e^{\mu_B+\mu_{A_1B}+\mu_{A_2B}+\mu_{A_3B}+\mu_{A_1A_2A_B}})
\end{cases}$$

After describing the components of the model in more detail, giving intuition behind the model,our task now is to generalize the definition of log-linear model to model the conditional probability.

Definition 3.2
- *X : set of possible children in the network.*
- *Y : set of possible parents in the network.*
- *f :X*Y → ℜ : function maps interactions between variables of the network, with*

$$f(x, y_1, ..., y_n) = \log P(x, y_1, ..., y_n)$$

For any x ∈ X and y ∈ Y, the model defines a conditional probability :

$$P(x|y_1, ..., y_n) = \frac{e^{f(x,y_1,...,y_n)}}{\sum_{y \in Y} e^{f(y_1,...,y_n)}} \tag{3.8}$$

In the following part, we are focusing in links between these components and information provided by QPN.

3.1.2 Links between $\mu - terms$ and semantic of QPN

We consider the first example, Figure 3.1. If we suppose that $p_0 = P(B = 0|A = 0)$ and $p_1 = P(B = 0|A = 1)$, we obtain :

$$\mu_B = \log \frac{1}{0.4} = \log \frac{1 - p_0}{p_0}$$
$$\mu_{AB} = \log \frac{1}{0.2} - \log \frac{1}{0.4} = \log \frac{1 - p_1}{p_1} - \log \frac{1 - p_0}{p_0}$$

In our QPN we have the node A affect positively the node B so :

$$p_0 > p_1 \Rightarrow p_0 = p_1 + \delta, \quad \text{with} \quad \delta > 0$$
$$e^{\mu_{AB}} = \frac{1 - p_1}{p_1} \times \frac{1 - p_1 - \delta}{p_1 + \delta} > 1 \Rightarrow \mu_{AB} > 0$$

If we suppose that $p_0 = p_1 \Rightarrow \mu_{AB} = 0$
The positive influence of QPN helps us to determine the sign of μ_{AB}. If A an B are independents so $\mu_{AB} = 0$, then we can remove this μ-term from our log-linear model.
We consider the second example, Figure 3.2. After calculating the values of μ_{A_1B} and μ_{A_2B} we find $\mu_{A_1B} > 0$ and $\mu_{A_2B} > 0$ which confirms our proposal in the first example : the positive influences A_1B and A_2B tell us about the signs of the interaction terms μ_{A_1B} and μ_{A_2B}. Thus, an interaction term of order 1 between two nodes (parent-child) is a qualitative influence in a QPN.
With probabilities given in the table of probabilities of Figure 3.2 we find $\mu_{A_1A_2B} \neq 0$ which is explained by the existence of a product synergy between A_1 and A_2 according to the value of B. Thus, the term $\mu_{A_1A_2B}$ corresponds precisely to a product synergy. This result confirms our proposal in the example 3 illustrated the first part when we simplified the saturated log-linear model by deleting terms of order 2 corresponding to the synergies that do not appear in our QPN.
To determine the value of $\mu_{A_1A_2B}$ we demonstrated that we need an information about the table $P(B|A1, A2)$.
We suppose now that $\mu_{A_1A_2B} = 0$, if $P(B = 0|A1 = 1, A2 = 1) = \alpha \Rightarrow \mu_{A_1A_2B} = log\frac{1-\alpha}{\alpha} - \mu_{A_1B} - \mu_{A_2B} - \mu_B \Rightarrow \alpha = \frac{14}{15}$ and $1 - \alpha = \frac{1}{15}$

QPN	Log linear model
Qualitative influence	μ_{A_iB}
Positive influence	$\mu_{A_iB} > 0$
Negative influence	$\mu_{A_iB} < 0$
Product synergy	$\mu_{A_iA_jB}$
lack of synergy	$\mu_{A_iA_jB} = 0$
presence of synergy	$\mu_{A_iA_jB} \neq 0$

TABLE 3.1 – *Links between semantics of QPN and $\mu -$ terms*

We summarize in Table 2 links between the semantics of QPN and μ-terms of log-linear model.

Conclusion

Summary

This paper has two main goals. In our first contribution, we give the main correspondences that we made between the CM, QPNs, and CBNs in order to propose a global architecture CMQPNCBN. In order to facilitate modeling process and making inferences, we propose in this paper to take advantage benefits of the three models, i.e., CMs for the elicitation process, QPNs for efficient visualization and CBNs for the reasoning process. We propose an incremental method that consists to firstly build the CM and then transform it into a QPN which combines causal modeling techniques and probability theory. The structure of the obtained QPN corresponds to the one of the constructed CM from experts with some modification regarding the direct, indirect relations, circular relation and dependency between variables. Secondly, it consists to use observational data with information from our QPN already built to construct our CBN. This method combines the strengths of causal relationships and Bayesian networks and reduces the limitations of both. In order to improve our process construction we propose a complementary direction concerns an enrichment process by the integration of semantic knowledge to anticipate construction and visualization tasks.
We illustrated the proposed approach on an example of decision problem which concerns the analysis of the various factors that may influence the risk of an accident occurring on a road.

In our second contribution, we describe how to obtain the probabilities table of causal a Bayesian network suitable for qualitative influences, product and additive synergies viewed on a QPN. To handle this objective, we begin by studying the log-linear models and make links between terms of this model and the CPTs. We have thus given a general definition of the log-linear model well suited to a conditional probability. Then, we highlight the relationships between the interaction terms of log-linear model and semantics of the QPN.

We conclude that resolving log-linear model needs to return to the experts to give some probabilities. To obtain a solvable equations system we add constraints on the interaction terms from highest order terms to lowest order terms. In practice, it is difficult to take account interaction terms of order greater than or equal to 3. Consequently, we proposed to be limited to terms of order 2 because in a QPN we have qualitative influences that involve two variables and that correspond to terms of order 1 and synergies that involve three variables and that correspond to terms of order 2.

Last, even if this second goal is not yet finished, we can see the work presented in this paper as a first step about learning causal Bayesian networks using log-linear model, when the information is expressed by a user with a qualitative probabilistic network.

Issues for Future Research

Our framework offers several opportunities for future research. Indeed, throughout our process we unveiled several communication problems between different structures that are not yet solved these problems give rise to future works, among them the transformation of CPTs constructed by a Bayesian approach to causal statements with strength impact. Also, the expansion for better interactions with the ontology axioms during the construction process, the use of other types of semantic relations to improve the construction process and the visualization task.

Now, we are carrying out research for the generalization of an automatic method based on log-linear model to learn causal Bayesian networks from information expressed by a user with a qualitative probabilistic network. This task is rather complicated since we would like to allow the program to calculate the synergistic CPTs of any combination of parent nodes with different information type that can be provided such as ; values of strength impact, and probabilistic intervals,

Annexes

A

Sommaire

A.1 The d-separation criterion 62
A.2 The representation hypothesis 63

A.1 The d-separation criterion

The criterion of d-separation allows us to study how information flows in a graph. Let us consider three variables X,Y and Z, which are represented as nodes in a DAG G. independence in A independent $C|B$ means that knowing the value of B blocks information about C being relevant to A, and vice versa. That is to say Z "blocks" all paths from node X to node Y. By blocking we mean stopping the flow of information. These concepts apply not only between three nodes, but also between sets of nodes. More generally, Pearl defined d-separation as follow [reference 85] :

Definition d-separation

A path p is said to be d-separated by a subset of node Z if and only if :

i) p contains serial or diverging connection such that the middle node is in Z , or

ii) p contains a converging connection such that the middle node is not in Z and no descendant of that node is in Z.

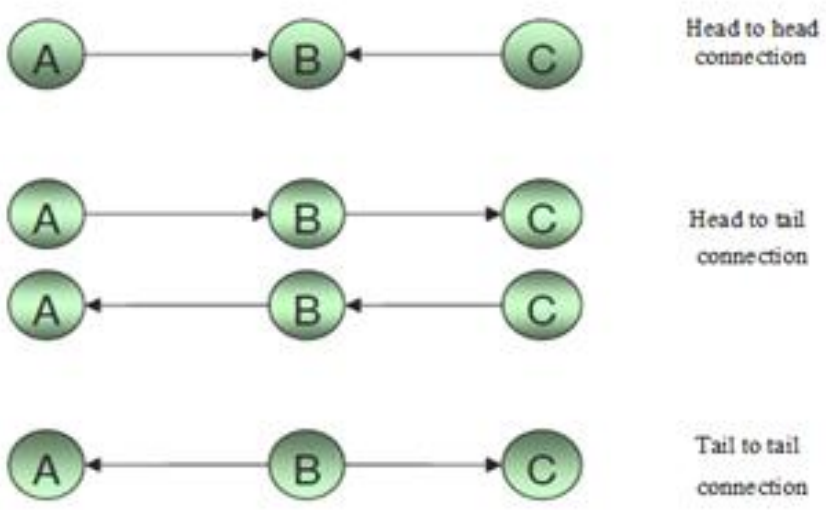


Figure A.1 – *Flow of information.*

The Markov equivalence

It's a fundamental property in learning BNs. Two BN structures G1 and G2 are said to be equivalent, if they can be used to represent the same set of probability distributions. More formally, Chickering [référence18] defines the Markov equivalence as follows :

Definition A.1 *Two DAGs G1 and G2 are equivalent if for every BN1=(G1, teta1), there exists a BN2=(G2, teta2) such that BN1 and BN2 define the same probability distribution, and vice versa.*

Moreover, Verma and Pearl propose the following definition which provides a graphical criterion for determining the equivalence of two DAGs :

Theorem A.1 *Two DAGs are equivalent if and only if they have the same skeletons and the same v-structures.*

The use of Bayesian networks is almost always in reference to an interpretation that is to say, more precisely, an interpretation of the arrows that appear in the graphs that make up the Bayesian networks. The interpretations are numerous. In our study we focus on the interpretation of causal Bayesian networks.

A.2 The representation hypothesis

1. Causal relata can be represented by variables ;
2. Direct causation among the variables in a given set can be represented by a binary relation ; this corresponds arrow (and especially an arrow of a Bayesian network) connects two points.
3. The graph representing direct causation among the variables in a given set is acyclic. This third representation hypothesis extends the second. The asymmetry of the binary relation which represents the causal Bayesian network running from this that the graphs are oriented Bayesian and acyclic.

Faithfulness assumption : Let G be a directed acyclic graph over a set of variables V, and p a probability distribution over V. (G ; p) satisfies Faithfulness if there are no probabilistic independences for p other than the ones stemming from (G ; p) being a Bayesian network. The faithfulness assumption allows us to move from probability distribution to graph.

The causal Markov condition : in a set of variables, a given variable is probabilistically independent of all its non-effects conditional on its direct causes.

Causal sufficiency assumption : This assumption is satisfied if there exist no common unmeasured (also known as hidden or latent) variables in the domain that are influencing one or more observed variables of the domain.

Bibliographie

F. Allard-Poesi. Cartes cognitives : pour ne pas jeter le bébé avec l'eau du bain. Dans *Ve Conférence Internationale de Management Stratégique*, 1996. (Cité page 8.)

R. Axelrod. Structure of decision : the cognitive maps of political elites. *Princeton, N.J.*, 1976. (Cité pages 6, 7, 8, 10, 11 et 12.)

G.A. Banini et R.A. Bearman. Application of fuzzy cognitive maps to factors affecting slurry rheology. *International Journal of Mineral Processing*, 52 :233–244, 1998. (Cité page 12.)

M. Ben Messaoud. *SemCaDo : an approach for serendipitous causal discovery and ontology evolution*. PhD thesis, Ecole Polytechnique de l'Université de Nantes, 2012. (Cité page 35.)

M. Ben Messaoud, P. Leray, et N. Ben Amor. Integrating ontological knowledge for iterative causal discovery. *In Proceedings of the 10th European Con ference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Verona, Italy*, pages 168–179, 2009. (Cité page 21.)

Y.M. Bishop, S.E. Fienberg, et P.W. Holland. *Discrete multivariate analysis*. MIT Press, Cambridge, Mass. [u.a.], 1975. ISBN 0262021137. (Cité page 51.)

H. Borchani, M. Chaouachi, et N. Ben Amor. Learning causal bayesian networks from incomplete observational data and interventions. Dans *In proceedings of Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 2007. (Cité page 21.)

M. Bougon. Uncovering cognitive maps : The selfq technique. *in Beyond method : strategies for social research, G. Morgan (ed.)*, pages 112–144, 1983. (Cité page 11.)

M. Bougon, K.E. Weick, et D. Binkhorst. Cognition in organizations : An analysis of the utrecht jazz orchestra. *Administrative Science Quarterly*, 22(4), 1977. (Cité pages 7 et 8.)

J. Carvalho et J. Tomé. Rule based fuzzy cognitive maps - qualitative systems dynamics. *In Proceedings of the 19th International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2000)*, pages 407–411, 1999. (Cité page 12.)

R.N. Carvalho. *Probabilistic ontology : Representation and modeling methodology*. PhD thesis, George Mason University, 2011. (Cité page 35.)

B. Chaib-draa. Causal maps : Theory, implementation, and practical applications in multiagent environments, 2002. (Cité pages 6 et 12.)

L. Chauvin, D. Genest, et S. Le Dorze, A.and Loiseau. Cartes cognitives : une exploitation à base d'échelle, vue et profil. Dans Ali Khenchaf et Pascal Poncelet, éditeurs, *EGC*, volume RNTI-E-20 de *Revue des Nouvelles Technologies de l'Information*, pages 503–514. Hermann-Éditions, 2011. (Cité pages 7, 34 et 38.)

R. Christensen. *Log-Linear Models and Logistic Regression*. Springer, 1997. (Cité page 51.)

G.F. Cooper et T. Dietterich. A bayesian method for the induction of probabilistic networks from data. *In Machine Learning*, pages 309–347, 1992. (Cité page 20.)

F. Corset. *Aide à l'optimisation de maintenance à partir de réseaux Bayésiens et fiabilité dans un contexte doublement censuré*. PhD thesis, Université Joseph Fourier, 2003. (Cité page 52.)

A. Devitt, B. Danev, et K. Matusikova. Constructing bayesian networks automatically using ontologies. Dans *Second Workshop on Formal Ontologies Meets Industry (FOMI 2006)*, Applied Ontology 0 (2006), Trento, Italy, 2006. IOS Press. (Cité page 35.)

J.A. Dickerson et B. Kosko. Virtual worlds as fuzzy cognitive maps. *Presence*, 3(2) :73–89, 1994. (Cité page 6.)

Z. Ding et Y. Peng. A probabilistic extension to ontology language OWL. Dans *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004. (Cité page 34.)

T. Doan, P. Haddawy, T. Nguyen, et D. Seetharam. A hybrid bayesian network modeling environment. Dans *the National Computer Science and Engineering Conference (NCSEC)*, Bangkok, Thailand,, 1999. (Cité page 34.)

M.J. Druzdzel et M. Henrion. Efficient reasoning in qualitative probabilistic networks. Dans *In Proceedings of the 11th National Conference on Artificial Intelligence (AAAI–93*, pages 548–553, 1993. (Cité page 26.)

F. Eberhardt, C. Glymour, et R. Scheines. On the number of experiments su-cient and in the worst case necessary to identify all causal relations among n variables. *In UAI Press*, pages 178–184, 2005. (Cité page 21.)

S. Fenz, A.M. Tjoa, et M. Hudec. Ontology-based generation of bayesian networks. Dans Leonard Barolli, Fatos Xhafa, et Hui-Huang Hsu, éditeurs, *CISIS*, pages 712–717. IEEE Computer Society, 2009. (Cité page 35.)

D. Genest et S. Loiseau. Modélisation, classification et propagation dans des réseaux d'influence. *Technique et Science Informatiques*, 26(3-4) :471–496, 2007. (Cité page 10.)

T.R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. *Int. J. Hum.-Comput. Stud.*, 43 :907–928, December 1995. (Cité page 29.)

Y.B. He et Z. Geng. Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9 :2523–2547, 2008. (Cité page 21.)

D. Hume. A treatise of human nature (i) of the understanding, volume 1 of history of economic thought books. *McMaster University Archive for the History of Economic Thought*, 1739. (Cité page 8.)

P. Judea. Comment : Graphical models, causality and intervention. *Statistical Science*, 8(3) :266–269, 1993. (Cité page 17.)

P. Judea. *Causality : Models, reasoning and inference*. 2000. (Cité pages 13, 14 et 16.)

J.J. Jung, K.Y. Jung, et G. Jo. Ontological cognitive map for sharing knowledge between heterogeneous businesses. Dans Adnan Yazici et Cevat Sener, éditeurs, *Computer and Information Sciences - ISCIS 2003, 18th International Symposium, Antalya, Turkey, November 3-5, 2003, Proceedings*, volume 2869 de *Lecture Notes in Computer Science*, pages 91–98. Springer, 2003. (Cité pages 33 et 34.)

B. Kosko. Fuzzy cognitive maps. *International Journal of Man-Machines Studies*, 26, 1986. (Cité pages 6, 8 et 12.)

K.B. Laskey et P.C.G. da Costa. Of starships and klingons : Bayesian logic for the 23rd century. Dans *UAI*, pages 346–353. AUAI Press, 2005. (Cité page 34.)

A. Le Dorze, L. Garcia, D. Genest, S. Loiseau, P . Leray, et B. Duval. Definition of cognitive maps where influences are interpreted as probabilities. Rapport technique, LERIA, Faculte des Sciences, Universite d'Angers LINA, Polytech Nantes, 2013. (Cité page 33.)

P.J.F. Lucas. Bayesian network modelling through qualitative patterns. *Artif. Intell.*, 163(2) :233–263, 2005. (Cité page 34.)

F. Mata Avila. Raisonnement qualitatif dans les systèmes multiagents basé sur les cartes causales. Master's thesis, Master's thesis, Université de Laval, Canada, August, 2002. (Cité page 12.)

S. Meganck, P. Leray, et B. Manderick. Learning causal bayesian networks from observations and experiments : A decision theoretic approach. *In Vicenç Torra, Yasuo Narukawa, Aïda Valls, and Josep Domingo-Ferrer, editors, MDAI, Springer*, 3885 of Lecture Notes in Computer Science :58–69., 2006. (Cité page 21.)

Y. Miao, C.K Siew, Z.Q Liu, et S. Li. cognitive network an extension of fuzzy cognitive map. Dans *In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 1999), page 43, Washington, DC, USA. IEEE Computer Society*, 1999. (Cité page 12.)

K.P. Murphy. Active learning of causal bayes net structure. Dans *Technical report, University of California, Berkeley, USA*, 2001. (Cité page 20.)

S. Nadkarni et P.P. Shenoy. A bayesian network approach to making inferences in causal maps. *European Journal of Operational Research*, 128(3) : 479–498, 2001. (Cité pages 32, 33 et 41.)

K.S. Park et S.H. Kim. Fuzzy cognitive maps considering time relationships. *Int. J. Hum.-Comput. Stud.*, 42(2) :157–168, 1995. (Cité page 12.)

D. Poignonec. *Apport de la combinaison cartographie cognitive ontologie dans la compréhension de la perception du fonctionnement d'un éosystème récifo-lagonaire de Nouvelle Calédonie par les acteurs locaux*. 2006. (Cité page 33.)

S. Renooij, S. Parsons, et P. Pardieck. Using kappas as indicators of strength in qualitative probabilistic networks. Dans *ECSQARU*, pages 87–99, 2003. (Cité pages 28 et 43.)

S. Renooij et L.C. van der Gaag. From qualitative to quantitative probabilistic networks. Dans *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, UAI'02, pages 422–429, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc. ISBN 1-55860-897-4. (Cité pages 43 et 45.)

K. Sedki et L. Bonneau De Beaufort. Cognitive Maps and Bayesian Networks for Knowledge Representation and Reasoning. Dans *IEEE 24th ICTAI*, pages 1035–1040, Grèce, 2012. (Cité page 33.)

P. Spirtes, C. Glymour, et R. Scheines. *Causation, Pre-diction and Search*. 2001. (Cité pages 14, 16 et 20.)

M.A. Styblinski et B.D. Meyer. Fuzzy cognitive maps, signal flow graphs, and qualitative circuit analysis. *Proceedings of the 2nd IEEE International Conference on Neural Networks (ICNN 88)*, 2 :549–556, 1988. (Cité page 12.)

R. Taber. nowledge processing with fuzzy cognitive maps. *Expert Systems With Application*, 2(1) :83–87, 1991. (Cité page 12.)

E. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55, 1948. (Cité page 6.)

S. Tong et D. Koller. Active learning for structure in bayesian networks. Dans *In International Joint Conference on Artificial Intelligence*, 2001. (Cité page 20.)

I. Truck. *Approche symbolique et floue des modificateurs linguistiques et leur lien avec l'aggregation*. PhD thesis, Université de Reims, France,, 2002. (Cité page 12.)

M. Uschold et M. Gruninger. Ontologies : Principles, methods and applications. *Knowledge Engineering Review*, 11 :93–136, 1996. (Cité pages 30 et 31.)

F. Van Kouwen, C. Dieperink, P Schot, et M. Wassen. Computer-supported cognitive mapping for participatory problem structuring. *Environment and Planning A*, 41(1) :63–81, 2009. (Cité pages 33 et 41.)

F. van Kouwen, S. Renooij, et P. Schot. Inference in qualitative probabilistic networks revisited. *Int. J. Approx. Reasoning*, 50(5) :708–720, Mai 2009. ISSN 0888-613X. (Cité pages 26, 27 et 45.)

M.P. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44(3) :257–303, 1990. (Cité pages 6 et 22.)

J. Woodward. Making things happen : A theory of causal explanation. Dans *Oxford University Press, Oxford*, 2003. (Cité page 16.)

Y. Yang et J. Calmet. OntoBayes : An Ontology-Driven Uncertainty Model. Dans *CIMCA '05 : Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce Vol-1 (CIMCA-IAWTIC'06)*, pages 457–463, Washington, DC, USA, 2005. IEEE Computer Society. (Cité page 35.)

J.Y. Zhang, Z.Q. Liu, et S. Zhou. Quotient fcms-a decomposition theory for fuzzy cognitive maps. *Trans. Fuz Sys.*, 11(5) :593–604, Octobre 2003. ISSN 1063-6706. (Cité page 6.)