The Difference between Single-Valued and Multi-Valued Cases in the Compact Representation of CPD in Bayesian Networks

Qin Zhang

Automation School of Chongqing University, Chongqing 400044, P.R. China
zhangqin@cqu.edu.cn

Abstract. This paper addresses an important issue about the compact representation of the conditional probability distribution (CPD) applied in the well known Bayesian Networks in uncertain causality representation and probabilistic inference. That is, there is an essential difference between the single-valued cases and the multi-valued cases, while this difference does not exist when the CPD is represented in the conditional probability table (CPT). In other words, the present compact representation and inference methods applicable in the single-valued cases may not be applicable in the multi-valued cases as people usually think. A detailed example is provided to illustrate this problem. The solution is provided in the references by the author.

Keywords: knowledge representation, uncertainty, causality, probabilistic inference.

1 Introduction

It is well known that the typical representation of the conditional probability distribution (CPD) in the well known Bayesian Network (BN) is the conditional probability table (CPT). However, it is also noted that there are too many parameters to be specified in a CPT. For example, suppose a child variable has 5 parent variables and all the 6 variables have 5 states each, the number of conditional probabilities included in the CPT is 5^6=15625. They are too many for the users to specify. To provide the compact representations, many efforts have been made, such as noisy-OR [1], CSI [2], DCD [3], etc. However, it should be noted that many of the compact representations are presented for or illustrated with the binary variables, while actually these cases are single-valued but not multi-valued. The so called single-valued case is such a case in which only the causes of one state (denoted as the true state) of the child variable are specified. In contrast, the case in which the causes of more than one state of the child variable are specified separately is a multi-valued case. Note that the word "valued" indicates the child variable states whose causes are specified directly (not the complement of the other states). Since the binary child variables can be involved in either the single-valued cases or the multi-valued cases, it is not clear whether or not or how the present compact representation models applicable in the single-valued cases are

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 599 – 606, 2009.

also applicable in the multi-valued cases. Note that the binary single-valued cases and the binary multi-valued cases are different. It is explained below.

2 The Detailed Discussion on the Essential Difference

Usually, people need only specify the causes of the true state of a binary variable X_n. Suppose $X_{n,1}$ denotes the true state and $X_{n,2}$ denotes the false state[1]. The reason why I use $X_{n,1}$ instead of X_n to denote the true state and use $X_{n,2}$ instead of $\overline{X}_n$ to denote the false state is for convenience of indicating the difference between the binary multi-valued cases and the binary single-valued cases. In the multi-valued cases, the states of a variable are in the identical positions. For the well known burglary (X_1), earthquake (X_2) and alarm (X_3) example in noisy-OR model [1], $X_{3,1}$ can be caused by either $X_{1,1}$ or $X_{2,1}$ independently. It is easy for the domain experts to give the special conditional probabilities (will be explained later) of $X_{3,1}$ caused by $X_{1,1}$ and $X_{2,1}$ independently, while it is not easy for them to give the CPT directly. This is because the burglary and earthquake are totally different domains. However, it should be noted that in this example, only the causes of $X_{3,1}$ are specified, while the causes of $X_{3,2}$ must *not* be specified, because $X_{3,2}$ has already been given as the complement of $X_{3,1}$. Figure 1 illustrates this binary single-valued case, in which, ⟶ represents the state level causal link.

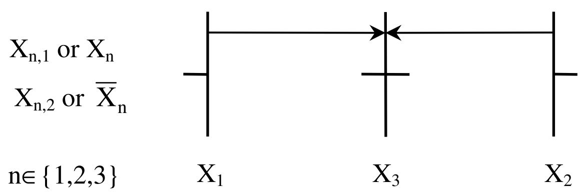


Fig. 1. The illustration for the binary single-valued case

However, the real world is not always so simple. For the example of a simple memory circuit, it has two identical states/outputs: 0 or 1. This is a typical binary variable. Similar to the alarm variable, we may denote this variable as X_3 with $X_{3,1}$ representing state "0" and $X_{3,2}$ representing state "1". But differently, both $X_{3,1}$ and $X_{3,2}$ can be caused by different events. Suppose event $X_{1,1}$ causes $X_{3,1}$ and $X_{2,1}$ causes $X_{3,2}$, with independently given special conditional probabilities $p_{3,1;1,1}$ and $p_{3,2;2,1}$ respectively, where the subscripts nk;ij denotes that event X_{nk} is caused by event X_{ij}.

The reason why the word "special" is put in front of "conditional probabilities" is because usually $p_{nk;ij}\neq\Pr\{X_{nk}|X_{ij}\}$. In fact, $p_{nk;ij}$ is the probability of the linkage event

[1] In this paper, X_{nk} denotes either the kth state/value of the variable X_n or the event that X_{nk} is true. The upper case letters denote variables or events. The lower case letters denote the probabilities of events, e.g., $x_{nk}\equiv\Pr\{X_{nk}\}$. The difference between the variable X_n and the event X_{nk} is that X_{nk} has two subscripts, in which the second subscript denotes the state or indexes the specific value of X_n.

in DCD and is the probability of the inhibitor in noisy-OR [1]. Similar notations are also used in [4] in which $p_{3,1;1,1}$ and $p_{3,2;2,1}$ are denoted as $c_{X1,1}(X_{3,1})$ and $c_{X2,1}(X_{3,2})$, so that the two types of conditional probabilities are distinctive.

Another example of multi-valued cases is the sex (X_3) that has two identical states: "male" ($X_{3,1}$) or "female" ($X_{3,2}$). The biological causes of the two states may be different and specified separately.

Abstractly, a simple binary multi-valued case is illustrated in figure 2. The difference of figure 2 from figure 1 is that in figure 2, the causes of both $X_{3,1}$ and $X_{3,2}$ are specified separately.

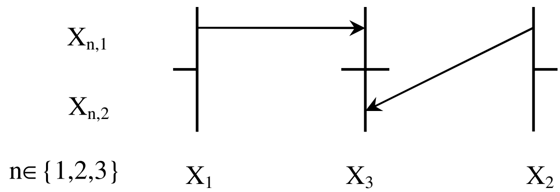


Fig. 2. The illustration for the binary multi-valued case

The multi-valued cases are often encountered when the child variable has more than two states. For the example of a temperature variable (X_3), it may have "normal" ($X_{3,1}$), "high" ($X_{3,2}$), "very high" ($X_{3,3}$), "low" ($X_{3,4}$) and "very low" ($X_{3,5}$) five states. The causes of them are usually different and should be specified separately.

It is well known that the probabilities of all states of a variable must sum up to 1 in any case. This can be called the *normalization*. The single-valued cases always satisfy the normalization, because the conditional probability of the true state Pr{true|condition} cannot be greater than 1 and the conditional probability of the false state Pr{false|condition} is just the complement of the true state, i.e., Pr{false|condition}=1−Pr{true|condition}. However, in the multi-valued cases, the situation is different. To illustrate this, suppose that in figure 2, $X_{1,1}$ causes $X_{3,1}$ with $p_{3,1;1,1}$=0.7 and $X_{2,1}$ causes $X_{3,2}$ with $p_{3,2;2,1}$=0.8. In this situation, if we calculate the conditional probabilities of $X_{3,1}$ and $X_{3,2}$ as in the single-valued case separately, the normalization is not satisfied, because, as two identical states, $\Pr\{X_{3,1}|X_{1,1}X_{2,1}\}=p_{3,1;1,1}$ and $\Pr\{X_{3,2}|X_{1,1}X_{2,1}\}=p_{3,2;2,1}$ separately, while

$$\Pr\{X_{3,1}|X_{1,1}X_{2,1}\}+\Pr\{X_{3,2}|X_{1,1}X_{2,1}\}=p_{3,1;1,1}+p_{3,2;2,1}=0.7+0.8=1.5>1$$

It seems that equation (1) always satisfies the normalization in any multi-valued case:

$$\Pr\{X_{nk} \mid E\} = \frac{\Pr\{X_{nk}E\}}{\sum_k \Pr\{X_{nk}E\}} . \tag{1}$$

In which, E represents the evidence or condition. This equation is generally used to achieve the normalization of the multi-valued cases. For example, in [5], we find the following words:

"In this and subsequent examples, we assume that variables are Boolean (i.e., with domain {*true, false*})"; "The theory and the implementations are not restricted to binary variables".

We also find the following equations in [5]:

$$P(X \mid E_1 = o_1 \wedge \cdots \wedge E_s = o_s) = \frac{P(X \wedge E_1 = o_1 \wedge \cdots \wedge E_s = o_s)}{P(E_1 = o_1 \wedge \cdots \wedge E_s = o_s)};$$

$$P(E_1 = o_1 \wedge \cdots \wedge E_s = o_s) = \sum_{v \in dom(X)} P(X = v \wedge E_1 = o_1 \wedge \cdots \wedge E_s = o_s);$$

$$P(X = v_i \mid e) = \frac{P(X = v_i \wedge e)}{\sum_{v_j} P(X = v_j \wedge e)}.$$

Where $X_i=O_i$ denotes a piece of evidence, $X=v_j$ denotes the event of a hypothesis and $v \in dom(X)$ denotes all the possible states/values that X can have.

These equations are exactly the same as equation (1). It seems that many researchers think that the single-valued cases are generally the same as the multi-valued cases, except that equation (1) should be used to satisfy the normalization. However, this is incorrect, because equation (1) is valid only when $\sum_k \Pr\{X_{nk} \mid E\} = 1$. This can be seen in equation (2):

$$\Pr\{X_{nk} \mid E\} = \frac{\Pr\{X_{nk}E\}}{\Pr\{E\}} = \frac{\Pr\{X_{nk}E\}}{\Pr\{E\}\sum_k \Pr\{X_{nk} \mid E\}} = \frac{\Pr\{X_{nk}E\}}{\sum_k \Pr\{X_{nk}E\}}. \tag{2}$$

In the case of using CPT to represent CPD, $\sum_k \Pr\{X_{nk} \mid E\} = 1$ is always satisfied, because E is a state combination of the conditional variables and the normalization of X_{nk} is always satisfied in the CPT for the given E. However, in the case of the compact representation, the situation is different. If we use the special conditional probabilities to calculate $\Pr\{X_{nk}|E\}$ separately as in the single-valued cases, $\sum_k \Pr\{X_{nk} \mid E\} = 1$ is usually not satisfied. The above example has shown this. Therefore, the compact representations and inference algorithms applicable in single-valued cases cannot be automatically applied in multi-valued cases.

To illustrate this in details, let E in equation (1) be $E_j=SCPV_{n;j}$, where $SCPV_{n;j}$ denotes the state combination #j of the parent variables of X_n, i.e., $SCPV_{n;j} = \bigcap_i X_{ik_{ij}}$, k_{ij} indexes the state of X_i included in $SCPV_{n;j}$. Then, equation (1) can be written as

$$\Pr\{X_{nk} \mid E_j\} = \frac{\Pr\{X_{nk}E_j\}}{\sum_k \Pr\{X_{nk}E_j\}} = \frac{\Pr\{X_{nk} \mid E_j\}}{\sum_k \Pr\{X_{nk} \mid E_j\}} = \alpha_{n;j}\Pr\{X_{nk} \mid E_j\}, \tag{3}$$

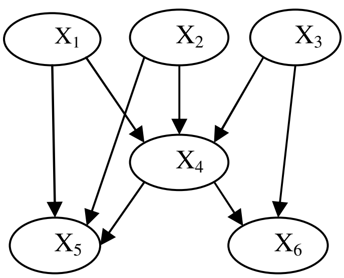
$$\alpha_{n;j} \equiv 1/\sum_{k} \Pr\{X_{nk} \mid E_j\}\,. \tag{4}$$

Where $\alpha_{n;j}$ is the normalization factor, so that equation (3) sums up to 1 with respect to k.

It is obvious that equation (1) is equivalent to equations (3) and (4).

In both equations (3) and (4), the $\Pr\{X_{nk}|E_j\}$ on the right side is the separately calculated conditional probability and the $\Pr\{X_{nk}|E_j\}$ on the left side of equation (3) is the normalized conditional probability. Note that in equation (4), the normalization factor $\alpha_{n;j}$ is not a constant but a variable depending on E_j=$SCPV_{n;j}$.

For the example shown in figure 3, which is the figure 1 in [2], suppose all variables are binary and only the causes of $X_{4,1}$ are specified with CSI. The case is single-valued, because $\Pr\{X_{4,2}|E_j\}=1-\Pr\{X_{4,1}|E_j\}$. Suppose we additionally specify the causes of $X_{4,2}$ separately as shown in figure 4, the case becomes multi-valued. Note that X_1 is not a parent variable of $X_{4,2}$, while X_1 is a parent variable of $X_{4,1}$. This is allowed in the separate specifications in the multi-valued cases.



j	E_j=$SCPV_4$	$\Pr\{X_{4,1}|E_j$
1	$X_{1,1}X_{2,1}X_{3,1}$	p_1=0.2
2	$X_{1,1}X_{2,1}X_{3,2}$	p_1=0.2
3	$X_{1,1}X_{2,2}X_{3,1}$	p_1=0.2
4	$X_{1,1}X_{2,2}X_{3,2}$	p_1=0.2
5	$X_{1,2}X_{2,1}X_{3,1}$	p_2=0.4
6	$X_{1,2}X_{2,1}X_{3,2}$	p_2=0.4
7	$X_{1,2}X_{2,2}X_{3,1}$	p_3=0.6
8	$X_{1,2}X_{2,2}X_{3,2}$	p_4=0.8

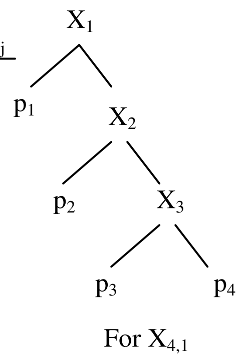


Fig. 3. The single-valued case with CSI

As being pointed out earlier, it is obvious that $\Pr\{X_{4,1}|E_j\}$ in figure 3 and $\Pr\{X_{4,2}|E_j\}$ in figure 4 cannot sum up to 1. If we insist on applying equations (3) and (4) to satisfy the normalization, we have to have

$\alpha_{4;1}=1/(\Pr\{X_{4,1}|E_1\}+\Pr\{X_{4,2}|E_1\})=1/(0.2+0.4)=1/0.6$
$\alpha_{4;2}=1/(\Pr\{X_{4,1}|E_2\}+\Pr\{X_{4,2}|E_2\})=1/(0.2+0.2)=1/0.4$
$\alpha_{4;3}=1/(\Pr\{X_{4,1}|E_3\}+\Pr\{X_{4,2}|E_3\})=1/(0.2+0.7)=1/0.9$
$\alpha_{4;4}=1/(\Pr\{X_{4,1}|E_4\}+\Pr\{X_{4,2}|E_4\})=1/(0.2+0.7)=1/0.9$
$\alpha_{4;5}=1/(\Pr\{X_{4,1}|E_5\}+\Pr\{X_{4,2}|E_5\})=1/(0.4+0.4)=1/0.8$
$\alpha_{4;6}=1/(\Pr\{X_{4,1}|E_6\}+\Pr\{X_{4,2}|E_6\})=1/(0.4+0.2)=1/0.6$
$\alpha_{4;7}=1/(\Pr\{X_{4,1}|E_7\}+\Pr\{X_{4,2}|E_7\})=1/(0.6+0.7)=1/1.3$
$\alpha_{4;8}=1/(\Pr\{X_{4,1}|E_8\}+\Pr\{X_{4,2}|E_8\})=1/(0.8+0.7)=1/1.5$

j	E_j=SCPV$_4$	Pr{$X_{4,2}$|E_j}
1	$X_{1,1}X_{2,1}X_{3,1}$	p_5=0.4
2	$X_{1,1}X_{2,1}X_{3,2}$	p_6=0.2
3	$X_{1,1}X_{2,2}X_{3,1}$	p_7=0.7
4	$X_{1,1}X_{2,2}X_{3,2}$	p_7=0.7
5	$X_{1,2}X_{2,1}X_{3,1}$	p_5=0.4
6	$X_{1,2}X_{2,1}X_{3,2}$	p_6=0.2
7	$X_{1,2}X_{2,2}X_{3,1}$	p_7=0.7
8	$X_{1,2}X_{2,2}X_{3,2}$	p_7=0.7

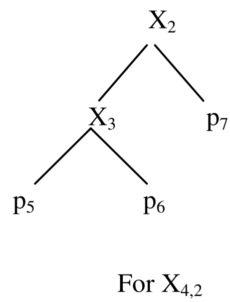


Fig. 4. The specification to Pr{$X_{4,2}$|E_j} with CSI in the multi-valued case

Then, from equation (3), the CPT can be calculated as shown in table 1.

Table 1. The CPT calculated from equation (3)

j	E_j=SCPV$_{4;j}$	Pr{$X_{4,1}$|E_j}	Pr{$X_{4,2}$|E_j}	$\alpha_{4;j}$
1	$X_{1,1}X_{2,1}X_{3,1}$	1/3	2/3	1/0.6
2	$X_{1,1}X_{2,1}X_{3,2}$	1/2	1/2	1/0.4
3	$X_{1,1}X_{2,2}X_{3,1}$	2/9	7/9	1/0.9
4	$X_{1,1}X_{2,2}X_{3,2}$	2/9	7/9	1/0.9
5	$X_{1,2}X_{2,1}X_{3,1}$	1/2	1/2	1/0.8
6	$X_{1,2}X_{2,1}X_{3,2}$	2/3	1/3	1/0.6
7	$X_{1,2}X_{2,2}X_{3,1}$	6/13	7/13	1/1.3
8	$X_{1,2}X_{2,2}X_{3,2}$	8/15	7/15	1/1.5

It is seen that the normalized CPT in table 1 is based on many different $\alpha_{4;j}$, $j\in\{1,\ldots,8\}$. Theoretically, the number of $\alpha_{n;j}$ equals to the number of E_j=SCPV$_{n;j}$, which means that the number of $\alpha_{n;j}$ can be huge. For the example of a child variable with five parent variables with five states each, the number of SCPV$_{n;j}$ is 5^5=3125. This is too many for domain experts to realize when they specify the causes and parameters of the states of X_n separately. The questions are: Why do we need so many different normalization factors? Are these different normalization factors realized by the domain experts when they specify the causes and parameters for the different multi-valued states separately? In other words, are these different normalization factors what the domain experts want? I do not think that these questions have always been clearly realized and answered when people apply equation (1).

Moreover, if we change the values of p_i as shown in table 2 (the new set of p_i), the normalized CPT remains same as in table 1. This is another issue that the domain experts may not realize. In fact, although the two sets of p_i work out a same CPT, they have different influences in the probability propagation through the causality chains

Table 2. The comparison between two sets of p_i

		The old set of p_i			The new set of p_i		
j	E_j=SCPV$_4$	$Pr\{X_{4,1}\|E_j\}$	$Pr\{X_{4,2}\|E_j\}$	$\alpha_{4;j}$	$Pr\{X_{4,1}\|E_j\}$	$Pr\{X_{4,2}\|E_j\}$	$\alpha_{4;j}$
1	$X_{1,1}X_{2,1}X_{3,1}$	$p_1=0.2$	$p_5=0.4$	1/0.6	$p_1=0.1$	$p_5=0.2$	1/0.3
2	$X_{1,1}X_{2,1}X_{3,2}$	$p_1=0.2$	$p_6=0.2$	1/0.4	$p_1=0.1$	$p_6=0.1$	1/0.2
3	$X_{1,1}X_{2,2}X_{3,1}$	$p_1=0.2$	$p_7=0.7$	1/0.9	$p_1=0.1$	$p_7=0.35$	1/0.45
4	$X_{1,1}X_{2,2}X_{3,2}$	$p_1=0.2$	$p_7=0.7$	1/0.9	$p_1=0.1$	$p_7=0.35$	1/0.45
5	$X_{1,2}X_{2,1}X_{3,1}$	$p_2=0.4$	$p_5=0.4$	1/0.8	$p_2=0.2$	$p_5=0.2$	1/0.4
6	$X_{1,2}X_{2,1}X_{3,2}$	$p_2=0.4$	$p_6=0.2$	1/0.6	$p_2=0.2$	$p_6=0.1$	1/0.3
7	$X_{1,2}X_{2,2}X_{3,1}$	$p_3=0.6$	$p_7=0.7$	1/1.3	$p_3=0.3$	$p_7=0.35$	1/0.65
8	$X_{1,2}X_{2,2}X_{3,2}$	$p_4=0.8$	$p_7=0.7$	1/1.5	$p_4=0.4$	$p_7=0.35$	1/0.75

when we apply the efficient inference algorithms based on the compact representations. To avoid this inconsistence, we may use the CPT in the probabilistic reasoning only. But this means that we give up the efficient algorithms applicable in the single-valued cases.

It should be pointed out that, in a single-valued case, the single-value is associated with only the child variable but *not* the parent variables. The number of *active* states of the parent variables can be more than one. Moreover, the single-valued child variable must be binary. More than two states of a single-valued child variable are meaningless, because the states not specified can be combined as one state. Figure 5 shows a single-valued case in which X_1 has three states and both X_1 and X_2 have two *active* states, while X_3 has only one *valued* state.

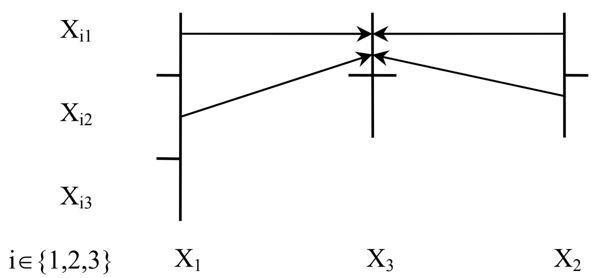


Fig. 5. Example of a single-valued case in which the parent variables have multiple active states

3 Conclusion

In conclusion, we should be careful whether or not equation (1) is what we want. In other words, the representations and algorithms applicable in the single-valued cases may not be applicable in the multi-valued cases when we apply the compact representation model in stead of CPT for representing CPD. The solution for compactly representing CPD and the corresponding inference algorithm has been presented by the author in [6-10].

References

[1] Pearl, J.: Probabilistic reasoning in intelligent systems. Morgan Kaufmann, San Mateo (1988)
[2] Boutilier, C., Friedman, N., Goldszmidt, M., Koller, D.: Context-specific independence in Bayesian Network. In: Proc. of UAI 1996 (1996)
[3] Zhang, Q.: Probabilistic reasoning based on dynamic causality trees/diagrams. Reliability Engineering and System Safety (46), 209–220 (1994)
[4] Poole, D., Zhang, N.L.: Exploiting contextual independence in probabilistic inference. Journal of Artificial Intelligence Research 18, 263–313 (2003)
[5] D'Ambrosio, B.: Local expression languages for probabilistic dependence. Int. J. Approximate Reasoning 139(1), 61–68 (1995)
[6] Zhang, Q.: DUCG - a new approach to deal with dynamical uncertain causality information - part I: the static discrete DAG case. Submitted for publication in IEEE Trans. System, Man and Cybernetics - Part A: Systems and Human
[7] Zhang, Q.: DUCG - a new approach to deal with dynamical uncertain causality information - part II: the dynamical DCG case. Submitted for publication in IEEE Trans. System, Man and Cybernetics - Part A: Systems and Human
[8] Zhang, Q.: DUCG - a new approach to deal with dynamical uncertain causality information - part III: the discrete dynamical case. Submitted for publication in IEEE Trans. System, Man and Cybernetics - Part A: Systems and Human
[9] Zhang, Q.: DUCG - a new approach to deal with dynamical uncertain causality information - part IV: The Continuous Variable and Uncertain Evidence. Submitted for publication in IEEE Trans. System, Man and Cybernetics - Part A: Systems and Human
[10] Zhang, Q.: An application of DUCG – The dynamical fault diagnoses and predictions of a nuclear power plant. Submitted for publication in IEEE Trans. System, Man and Cybernetics - Part A: Systems and Human