**Theorem of SIGMA-gate Inference in Bayesian Network**

Loc Nguyen

University of Science, Ho Chi Minh city, Vietnam

Tel: +84975250362 E-mail: ng\_phloc@yahoo.com

**Abstract**

Bayesian network is a powerful mathematical tool for doing diagnosis and assessment tasks. Parameter learning in Bayesian network is complicated study but I recognize that parameter learning becomes easy in some situations. Especially, when Bayesian network is weighted graph and its child node is aggregation of mutually independent parent nodes, there is a simple way to specify conditional probability tables which are parameters of Bayesian network. In this research, I propose and prove the theorem of SIGMA-gate inference which is the fundamental of such simple way helping us to transform weighted graph into Bayesian network. Note that the theorem is derived from works of authors Millán and Pérez-de-la-Cruz in their article “A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation” published in User Modeling and User-Adapted Interaction Journal on June 2002.

**Keywords:** Bayesian network, parameter learning, SIGMA-gate inference.

**1. Issued problem**

Bayesian network (BN) (Neapolitan, 2003, p. 40) is the directed acyclic graph (DAG) whose nodes are linked together by arcs or edges; each arc expresses the dependence relationships. There are many kinds of dependence relationships such as cause-effect, diagnosis, aggregation, prerequisite and etc. Note that we only study binary Bayesian network whose nodes are referred as *binary random variables* having two values 0 and 1. The strengths of dependences are quantified by conditional probability table (CPT). When one variable is conditionally dependent on another, there is a corresponding probability in CPT measuring the strength of such dependence; in other words, each CPT represents the local conditional probability distribution of a variable. If there is the arc from node *X* to *Y*, we call “*X* is parent of *Y*”or “*Y* is child of *X*”. Figure 1 is a typical example of Bayesian network, so-called wet-grass network (Murphy, 1998).



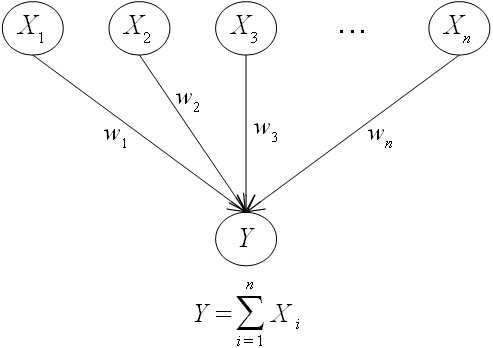
**Figure 1.** Wet-grass network

In above figure, event “cloudy” is cause of event “rain” or “sprinkler”, which in turn is cause of “grass is wet”. So we have three dependence relationships such as “1-cloudy to rain”, “2-rain to wet grass” and “3-sprinkler to wet grass”; moreover, each node is quantified by a CPT with note that the notation *P*(.) denotes the probability. It is easy to recognize that Bayesian network consists of a graph and a set of CPT (s) attached to nodes. Given a directed acyclic graph (DAG) whose arcs are weighted, there is a problem needs solved “how to determine CPT (s)”. This is the learning parameter problem when CPT (s) are considered as quantitative parameters. We narrow such problem into special DAG called aggregation graph in which child node is the aggregation of parent nodes; in other words, each arc expresses an aggregation relationship (Millán & Pérez-de-la-Cruz, 2002). If parent nodes are mutually independent, aggregation network becomes *sigma graph*. In other words, all parent nodes in sigma graph constitute a complete set of mutually exclusive random variables. Given sigma graph contains a set of parent nodes *X*1, *X*2,…, *Xn* and a child node *Y* where *Y* is the sigma sum of all *Xi* (s) and each arc *Xi* → *Y* is weighted.

Note that *Xi* and *Xj* are binary random variables and so we use notation to denote that *Xi* (s) are mutually independent. Hence, the *sigma sum* is interpreted that node *Y* is exclusive aggregation or exclusive union of nodes *Xi* (s) and so, the sigma sum sign ∑ does not express arithmetical addition. Binary variables *Xi* and *Y* represent events. Following are some notes about probabilistic events:

* The assignment “*Xi*=1” (“*Xi*=0”) means event *Xi* does (does not) occur.
* The assignment “*Y*=1” (“*Y*=0”) means event *Y* does (does not) occur.
* The probability that *Xi* does occur is denoted *P*(*Xi*=1). The probability that *Xi* does not occur is denoted *P*(*Xi*=0).
* The probability that *Y* does occur is denoted *P*(*Y*=1). The probability that *Y* does not occur is denoted *P*(*Y*=0).

Figure 2 depicts sigma graph.



**Figure 2.** Sigma graph

Now we prefer to use terms “sigma sum” instead of “exclusive aggregation” or “exclusive union” as usual. In sigma graph, parent node is called *partial node* or *source node* and child node is called *aggregative node* or *target node*. Note that statement “target node *Y* is aggregation of sources node *Xi* (s)” is the same to statement “each source node *Xi* is integrated into target node *Y*”.

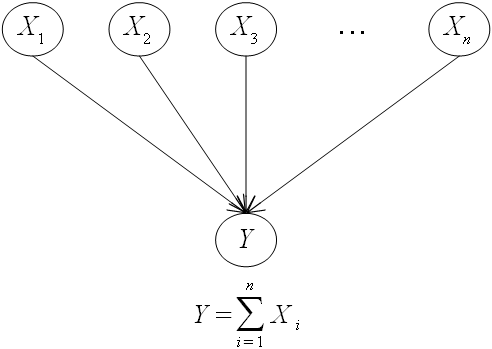
The main problem is how to transform sigma graph into Bayesian network; in other words, it is necessary to calculate all CPT (s) attached to nodes. Such problem is also called parameter learning. The next section mentions the theorem of SIGMA-gate inference which helps us to solve this problem.

**2. SIGMA-gate inference**

Suppose every node is binary, SIGMA-gate inference associated to sigma graph is based on three assumptions (Neapolitan, 2003, p. 157):

* *Aggregation inhibition*: Given an aggregation relationship denoted by arc *X*→*Y*, there is a factor *I* that inhibits *X* from integrated into *Y*. Factor *I* is called inhibition of *X*. That the inhibition *I* is turned off is the prerequisite of *X* integrated into *Y*.
* *Inhibition independence*: Inhibitions are mutually independent. For example, inhibition *I*1 of *X*1 is independent from inhibition *I*2 of *X*2.
* *Sigma condition*: Suppose we have a set of aggregation relationships in which *Y* is the aggregation of many sources *X*1, *X*2,… *Xn* and let *Ii* be the inhibition of *Xi*, the sigma condition states that “the target *Y* is the sigma sum of all sources *Xi* (s)”. So the sigma condition is merely the main aspect of sigma graph as aforementioned.

Concepts “aggregation inhibition”, “inhibition independence” and “sigma condition” are inspired from concepts “cause inhibition”, “exception independence” and “accountability” of noisy OR-gate model (Neapolitan, 2003, pp. 156-160) (Onisko, Druzdzel, & Wasyluk, 2001, pp. 168-170) in Bayesian network inference. Figure 3 also depicts sigma condition.



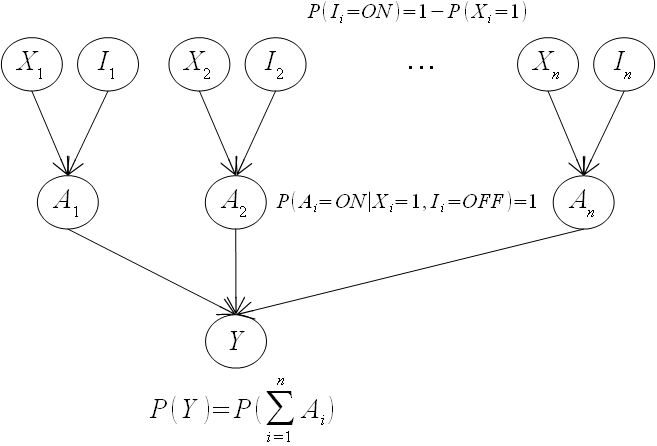
**Figure 3.** Sigma condition

Suppose we have *n* sources *X*1, *X*2,…, *Xn* and one target *Y*. According to “aggregation inhibition” and “inhibition independence” assumptions and let *Ii* be the inhibition of *Xi*. Let *Ai* be dummy variable so that *Ai* is *ON* (=1) if *Xi* is equal to 1 and *Ii* is *OFF* (=*0*), we have:

|  |
| --- |
| *P*(*Ai* =*ON* | *Xi* =1, *Ii* =*OFF*) = 1 |
| *P*(*Ai* =*ON* | *Xi* =1, *Ii* =*ON*) = 0 |
| *P*(*Ai* =*ON* | *Xi* =0, *Ii* =*OFF*) = 0 |
| *P*(*Ai* =*ON* | *Xi* =0, *Ii* =*ON*) = 0 |
|  |
| *P*(*Ai* = *OFF* | *Xi* =1, *Ii* =*OFF*) = 0 |
| *P*(*Ai* = *OFF* | *Xi* =1, *Ii* =*ON*) = 1 |
| *P*(*Ai* = *OFF* | *Xi* =0, *Ii* =*OFF*) = 1 |
| *P*(*Ai* = *OFF* | *Xi* =0, *Ii* =*ON*) = 1 |

Binary variables *Ai* (s) and *Ii* (s) also represent probabilistic events. According to sigma condition, the condition probability of *Y* is the probability of sigma sum of all *Ai* (s) where *Ai* (s) are mutually independent.

The sigma sum is interpreted that variable *Y* is exclusive aggregation or exclusive union of variables *Ai* (s). Figure 4 depicts SIGMA-gate model with regard to dummy variables *Ai* (s).



**Figure 4.** SIGMA-gate model

Now the strength of each aggregation relationship *Xi* → *Y* is quantified by the CPT *P*(*Y* | *Xi*). Suppose sources (*X*1, *X*2,…, *Xi*,…, *Xn*) become evidences having values (*x*1, *x*2,…, *xi*,…, *xn*). Let *P*(*Xi* =1) = *pi* be the probability of *Xi* = 1, the probability of inhibition of *Xi* is the inverse of *P*(*Xi* =1).

*P*(*Ii* =*ON*) = 1 – *P*(*Xi* =1) = 1 – *pi*

*P*(*Ii* =*OFF*) = *P*(*Xi* =1) = *pi*

Please pay attention that the set (*X*1, *X*2,…, *Xi*,…, *Xn*) is a partition of probability space; in other words, sum of all probabilities *P*(*Xi* =1) must be equal to 1.

Given the value set {*X*1=*x*1, *X*2=*x*2,…, *Xn*=*xn*}, let *K* be the set of *i* (s) such that *xi* = 1.

The goal of SIGMA-gate inference is to determine the posterior probability *P*(*Y* | *X*1, *X*2,…, *Xi*,…, *Xn*). We have:

(Due to SIGMA condition)

(Because *Ai* (s) are mutually independent)

(Because *Ai* is only dependent on *Xi*)

When sources {*X*1, *X*2,…, *Xn*} and target *Y* receive values, we have:

Applying the law of total probability with regard tothe set *K* of *i* (s) such that *xi* = 1, we have:

Where,

Suppose target event *Y* does not occur (*Y*=0), we have:

Recall that the set {*X*1, *X*2,…, *Xn*} is a partition of probability space, we have:

It implies

Where,

In conclusion, the theorem of SIGMA-gate inference states that *given target variable Y which is sigma sum of mutually independent source variables Xi* (*s*)*, the probability of Y is the sum of prior probabilities of Xi* (*s*) *which are equal to Y* as follows:

|  |  |
| --- | --- |
| Where, | (1) |

Please pay attention that the sum must be equal to 1 because the set (*X*1, *X*2,…, *Xi*,…, *Xn*) is a partition of probability space. Essentially, equation 1 is the same to equations specifying relationship between topics and subjects, invented by authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, pp. 292-295). However, equation 1 is more general and I prove it by different way. In other words, theorem of SIGMA-gate is only a simple and general case of works that were invented by authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002). I express my deep gratitude to authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002) for providing excellent works.

Going back issued problem with sigma graph in previous section when probabilities of source nodes *Xi* (s) are not specified; in other words, probabilities *P*(*Xi*) are not defined but each aggregation arc is weighted. Suppose *wi* is the weight of arc *Xi* → *Y* from node *Xi* to *Y*, the theorem of SIGMA-gate inference restates that *given target variable Y which is sigma sum of mutually independent source variables Xi* (*s*)*, the probability of Y is the sum of weights wi* (*s*) *with condition that the respective Xi* (*s*) *are equal to Y*.

|  |  |
| --- | --- |
| Where, | (2) |

Please pay attention that sum of all *wi* (s) must be equal to 1; in other words all weights *wi* (s) are normalized, . It is very easy to prove this variant of SIGMA-gate inference theorem. That the arc *Xi* → *Y* is weighted implies that the prior probability of (*Xi* = 1) is equal to *wi*.

Therefore, we have equation 2 as follows:

Where,

In general, the proof of SIGMA-gate inference is inspired from noisy OR-gate model (Neapolitan, 2003, p. 157) in Bayesian network. Authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002) invented the equation 2 by different way. Given aggregation relationships shown as sigma graph in figure 1, according to authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, p. 293), the conditional probability of *Y*=1 given random real variable *F* where 0 ≤ *f* ≤1 is considered as variable *F* itself.

Where *K* is the set of *i* (s) such that *Xi*=1. In learning context, each *Xi* represents a concept and *Y* represents a topic that is composed of many concepts.

Authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, p. 292) defined that *F* has distribution according to following equation:

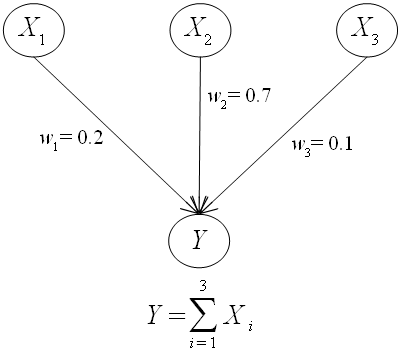
With regard to the equation above, authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, p. 292) claimed that if the total grade of given student is equal to the normalized sum of concept masteries, then, it is 100% confident that such student mastered these concepts, .

The probability of *Y*=1 is expectation of *F* due to total probability law as follows:

Authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, p. 293) determined equation 2 by calculating the expectation of *F* as follows:

According to the equation above, if student masters some concepts *Xi* (=1), the probability that she/he masters topic *Y* is equal to sum of weights of these concepts.

Now it is easy to transform the weighted sigma graph into Bayesian network. Given example where sigma graph has one target node *Y* and three source nodes *X*1, *X*2 and *X*3 whose weights are *w*1 = 0.2, *w*2 = 0.7 and *w*3 = 0.1, respectively. Figure 5 depicts given weighted sigma graph.



**Figure 5.** An example of sigma graph

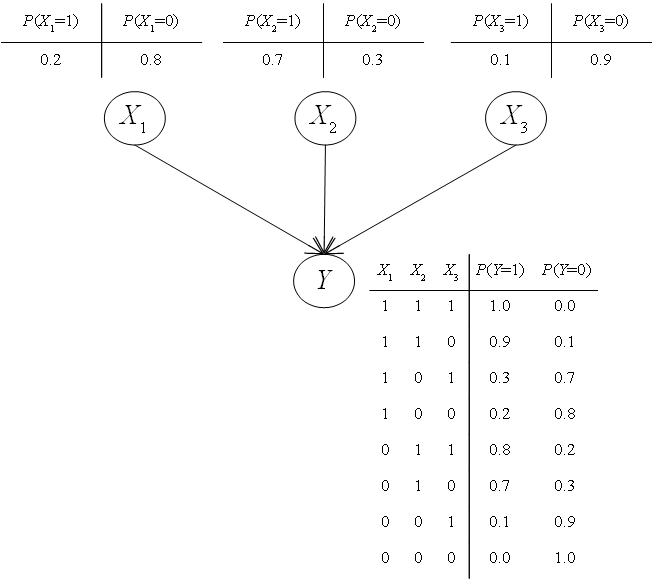
Applying theorem of SIGMA-gate inference, it is possible to determine prior probabilities of source nodes *X*1, *X*2 and *X*3 and conditional probability (CPT) of target node *Y* as below:

|  |  |
| --- | --- |
| *P*(*X*1=1)=*w*1=0.2 | *P*(*X*1=0)=0.8 |
| *P*(*X*2=1)=*w*2=0.7 | *P*(*X*2=0)=0.3 |
| *P*(*X*3=1)=*w*3=0.1 | *P*(*X*3=0)=0.9 |

|  |
| --- |
| *P*(*Y*1=1 | *X*1=1, *X*2=1, *X*3=1) = 1\*0.2 + 1\*0.7 + 1\*0.1 = 1.0 |
| *P*(*Y*1=1 | *X*1=1, *X*2=1, *X*3=0) = 1\*0.2 + 1\*0.7 + 0\*0.1 = 0.9 |
| *P*(*Y*1=1 | *X*1=1, *X*2=0, *X*3=1) = 1\*0.2 + 0\*0.7 + 1\*0.1 = 0.3 |
| *P*(*Y*1=1 | *X*1=1, *X*2=0, *X*3=0) = 1\*0.2 + 0\*0.7 + 0\*0.1 = 0.2 |
| *P*(*Y*1=1 | *X*1=0, *X*2=1, *X*3=1) = 0\*0.2 + 1\*0.7 + 1\*0.1 = 0.8 |
| *P*(*Y*1=1 | *X*1=0, *X*2=1, *X*3=0) = 0\*0.2 + 1\*0.7 + 0\*0.1 = 0.7 |
| *P*(*Y*1=1 | *X*1=0, *X*2=0, *X*3=1) = 0\*0.2 + 0\*0.7 + 1\*0.1 = 0.1 |
| *P*(*Y*1=1 | *X*1=0, *X*2=0, *X*3=0) = 0\*0.2 + 0\*0.7 + 0\*0.1 = 0.0 |

|  |
| --- |
| *P*(*Y*1=0 | *X*1=1, *X*2=1, *X*3=1) = 0\*0.2 + 0\*0.7 + 0\*0.1 = 0.0 |
| *P*(*Y*1=0 | *X*1=1, *X*2=1, *X*3=0) = 0\*0.2 + 0\*0.7 + 1\*0.1 = 0.1 |
| *P*(*Y*1=0 | *X*1=1, *X*2=0, *X*3=1) = 0\*0.2 + 1\*0.7 + 0\*0.1 = 0.7 |
| *P*(*Y*1=0 | *X*1=1, *X*2=0, *X*3=0) = 0\*0.2 + 1\*0.7 + 1\*0.1 = 0.8 |
| *P*(*Y*1=0 | *X*1=0, *X*2=1, *X*3=1) = 1\*0.2 + 0\*0.7 + 0\*0.1 = 0.2 |
| *P*(*Y*1=0 | *X*1=0, *X*2=1, *X*3=0) = 1\*0.2 + 0\*0.7 + 1\*0.1 = 0.3 |
| *P*(*Y*1=0 | *X*1=0, *X*2=0, *X*3=1) = 1\*0.2 + 1\*0.7 + 0\*0.1 = 0.9 |
| *P*(*Y*1=0 | *X*1=0, *X*2=0, *X*3=0) = 1\*0.2 + 1\*0.7 + 1\*0.1 = 1.0 |

Figure 6 depicts Bayesian network transformed from weighted sigma graph.



**Figure 6.** Bayesian network transformed from sigma graph

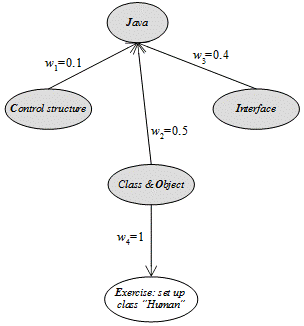
**3. An application of SIGMA-gate inference**

The main problem aforementioned is totally solved by SIGMA-inference when CPT (s) in Bayesian network are determined. Now we take account into a big application “how to design online course in order to teach students and assess their study results”. Suppose Java course is constituted of four concepts considered as hidden nodes whose links are aggregation relationships. Hidden nodes represent learning concepts such as “***J****ava*”, “***C****ontrol structure*”, “*Class &* ***O****bject*” and “***I****nterface*” with note that target node “***J****ava*” represents whole course. Additionally, there is an evidence node “***E****xercise: set up class Human*” which is an exercise. Evidence “***E****xercise: set up class Human*” proves whether or not she/he understands concept “*Class &* ***O****bject*”. All nodes and arcs constitute a graph with note that every node is binary random variable. The number in range [0, 1] that measures the relative importance of each relationship is defined by expert or teacher. In other words, this is the weight of arc from parent node to child node. All weights concerning the child variable will build up its CPT. Sum of weights of all arcs to each child node should be *1*. It means that each weight is normalized. These weights were proposed by authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002, p. 287), which express degrees of importance of concepts. Authors Millán and Pérez-de-la-Cruz claimed that “the weight of a concept can be computed as the number of days associated with it over the number of days associated with the topic it belongs to” (Millán & Pérez-de-la-Cruz, 2002, p. 288).

Your attention please, the relationship between hidden variable and evidence variable must be from hidden variable to evidence variable because the process that computes posterior probability of hidden variable with evidence is the knowledge diagnosis. So, evidence variable has no child and its parents must be hidden variables. In short, there are two kinds of relationships:

* Aggregation relationships among hidden variables. The aggregation relationship was mentioned in (Millán & Pérez-de-la-Cruz, 2002, p. 289). Note that the set of all parents of a hidden variable is the complete set of mutually exclusive hidden variables.
* Diagnostic relationships of hidden variables to evidences. The mastery of hidden concepts effects on the trust of evidences. However, if learner failed an examination, it is not sure about her/his lack of knowledge or ability because she/he can make a mistake unexpectedly.

Figure 7 depicts our course weighted graph representing Java course with note that such graph is the structure of Bayesian overlay model (Nguyen & Do, 2009). This structure follows the granularity hierarchy proposed in (Millán & Pérez-de-la-Cruz, 2002, pp. 287-288).



**Figure 7.** Weighted graph or overlay model

Note that evidence nodes are unshaded; otherwise, hidden nodes are shaded. Now it is necessary to specify CPT (s) of variables in Bayesian network (Bayesian overlay model).

It is easy to recognize that this weighted graph or overlay model is a composite sigma graph. What we need to do now is to transform overlay model into Bayesian network by applying SIGMA-gate inference. In this example, target node *J* (Java) has three source parents such as *C* (Control structure), *O* (Class & Object), and *I* (Interface) which in turn are corresponding to three weights of aggregation relationships such as *w*1=0.1, *w*2=0.5 and *w*3=0.4. These weights are also the prior probabilities of node *C*, *O* and *I*. We have:

|  |  |
| --- | --- |
| *P*(*C*=1) = *w*1 = 0.1 | *P*(*C*=0) = 0.9 |
| *P*(*O*=1) = *w*2 = 0.5 | *P*(*O*=0) = 0.5 |
| *P*(*I*=1) = *w*3 = 0.4 | *P*(*I*=0) = 0.6 |

It is easy to infer that target node *J* is the sigma sum of source nodes *C*, *O* and *I*. By applying SIGMA-gate inference, the conditional probabilities or CPT of node *J* is totally determined, for instance, we have:

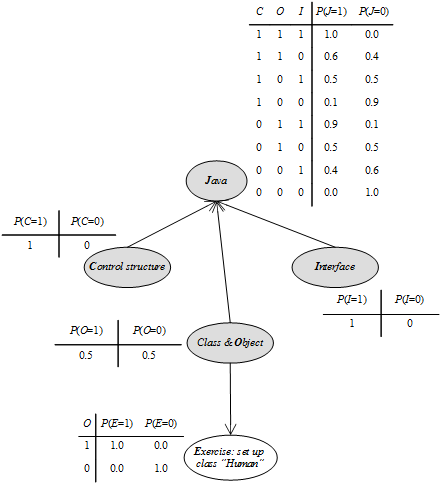
Following is the CPT of node *J*.

|  |
| --- |
| *P*(*J*=1 | *C*=1, *O*=1, *I*=1) = 1\*0.1 + 1\*0.5 + 1\*0.4 = 1.0 |
| *P*(*J*=1 | *C*=1, *O*=1, *I*=0) = 1\*0.1 + 1\*0.5 + 0\*0.4 = 0.6 |
| *P*(*J*=1 | *C*=1, *O*=0, *I*=1) = 1\*0.1 + 0\*0.5 + 1\*0.4 = 0.5 |
| *P*(*J*=1 | *C*=1, *O*=0, *I*=0) = 1\*0.1 + 0\*0.5 + 0\*0.4 = 0.1 |
| *P*(*J*=1 | *C*=0, *O*=1, *I*=1) = 0\*0.1 + 1\*0.5 + 1\*0.4 = 0.9 |
| *P*(*J*=1 | *C*=0, *O*=1, *I*=0) = 0\*0.1 + 1\*0.5 + 0\*0.4 = 0.5 |
| *P*(*J*=1 | *C*=0, *O*=0, *I*=1) = 0\*0.1 + 0\*0.5 + 1\*0.4 = 0.4 |
| *P*(*J*=1 | *C*=0, *O*=0, *I*=0) = 0\*0.1 + 0\*0.5 + 0\*0.4 = 0.0 |
|  |
| *P*(*J*=0 | *C*=1, *O*=1, *I*=1) = 0\*0.1 + 0\*0.5 + 0\*0.4 = 0.0 |
| *P*(*J*=0 | *C*=1, *O*=1, *I*=0) = 0\*0.1 + 0\*0.5 + 1\*0.4 = 0.4 |
| *P*(*J*=0 | *C*=1, *O*=0, *I*=1) = 0\*0.1 + 1\*0.5 + 0\*0.4 = 0.5 |
| *P*(*J*=0 | *C*=1, *O*=0, *I*=0) = 0\*0.1 + 1\*0.5 + 1\*0.4 = 0.9 |
| *P*(*J*=0 | *C*=0, *O*=1, *I*=1) = 1\*0.1 + 0\*0.5 + 0\*0.4 = 0.1 |
| *P*(*J*=0 | *C*=0, *O*=1, *I*=0) = 1\*0.1 + 0\*0.5 + 1\*0.4 = 0.5 |
| *P*(*J*=0 | *C*=0, *O*=0, *I*=1) = 1\*0.1 + 1\*0.5 + 0\*0.4 = 0.6 |
| *P*(*J*=0 | *C*=0, *O*=0, *I*=0) = 1\*0.1 + 1\*0.5 + 1\*0.4 = 1.0 |

Let *w*4 = 1 be weight of diagnostic relationships from hidden node *O* (Class & Object) to evidence nodes *E* (Exercise: set up class Human), conditional probabilities of *E* are totally determined as follows:

|  |  |
| --- | --- |
| *P*(*E*=1 | *O*=1) = 1\**w*4 = 1  *P*(*E*=1 | *O*=0) = 0\**w*4 = 0 | *P*(*E*=0 | *O*=1) = 0  *P*(*E*=0 | *O*=0) = 1 |

Hence, the Java course overlay model is transformed totally into Bayesian network as figure 8:



**Figure 8.** Bayesian overlay model and its parameters in full

**4. Conclusion**

The main aspect of this research is SIGMA-gate inference with attention that such inference is based on assumption of the mutual independence between source nodes which are integrated into only one target node. SIGMA-gate inference is inspired from OR-gate model in electronic circuit, Boolean algebra and Bayesian network inference (Neapolitan, 2003, pp. 156-160) (Onisko, Druzdzel, & Wasyluk, 2001, pp. 168-170), hence, the input events which are evidence and intermediate variables give out the results of diagnosis, assessments and etc. The inference is not only simple but also powerful method for designing tree-like Bayesian network such as learning course and diagnostic model. Note that the theorem of SIGMA-gate inference is a simple and general case of works invented by authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002).

**Acknowledgements**

I express my deep gratitude to authors Millán and Pérez-de-la-Cruz (Millán & Pérez-de-la-Cruz, 2002) for providing excellent works in their article “A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation” published in User Modeling and User-Adapted Interaction Journal on June 2002. The theorem of SIGMA-gate inference is derived from their works.

The first version of this research is the place to acknowledge Professor Nguyen, Minh-Man V. (Ho Chi Minh City University of Technology, Vietnam) who gave me an enthusiastic help to complete the theorem. The revised version expresses my deep gratitude to Professor Ho, Dang-Phuc (Vietnam Institute of Mathematics) who gave me valuable knowledge and motivation to enhance the research.

**Reference**

Millán, E., & Pérez-de-la-Cruz, J. L. (2002, June). A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation. (A. Kobsa, Ed.) *User Modeling and User-Adapted Interaction, 12*(2-3), 281-330. doi:10.1023/A:1015027822614

Murphy, K. P. (1998). *A Brief Introduction to Graphical Models and Bayesian Networks.* Retrieved 2008, from Kevin P. Murphy's home page: http://www.cs.ubc.ca/~murphyk/Bayes/bnintro.html

Neapolitan, R. E. (2003). *Learning Bayesian Networks.* Upper Saddle River, New Jersey, USA: Prentice Hall.

Nguyen, L., & Do, P. (2009). Combination of Bayesian Network and Overlay Model in User Modeling. *Proceedings of 4th International Conference on Interactive Mobile and Computer Aided Learning (IMCL 2009).* Amman, Jordan: Princess Sumaya University for Technology.

Onisko, A., Druzdzel, M. J., & Wasyluk, H. (2001). Learning Bayesian network parameters from small data sets: application of Noisy-OR gates. (T. Denoeux, Ed.) *International Journal of Approximate Reasoning, 27*(2), 165-182.