Hierarchical Inconsistent Qualitative Knowledge Integration for Quantitative Bayesian Inference

Rui Chang[1] **Martin Stetter**[2] **Wilfried Brauer**[1]

[1]Department of Computer Science, Technical University of Munich, Germany
[2]Information&Communication, Corporate Technology, Siemens AG, Munich, Germany

Abstract

We propose a novel framework for performing quantitative Bayesian inference based on qualitative knowledge. Here, we focus on the treatment in case of inconsistent qualitative knowledge. A hierarchical Bayesian model is proposed for integrating inconsistent qualitative knowledge by calculating a prior belief distribution based on a vector of knowledge features. Each inconsistent knowledge component uniquely defines a model class in the hyperspace. A set of constraints within each class is generated to describe the uncertainty in ground Bayesian model space. Quantitative Bayesian inference is approximated by model averaging with Monte Carlo methods. Our method is tested on ASIA network and results suggest that it enables reasonable quantitative Bayesian inference from a set of inconsistent qualitative knowledge.

Keywords: Qualitative knowledge modeling, Inconsistent knowledge integration, Bayesian networks, Bayesian inference, Monte Carlo simulation

1. Introduction

Bayesian models combine probability theory with graph theory and become increasingly important to design and analyze machine learning algorithms. Prior background knowledge can be combined with observed data to determine the probability distribution of a hypothesis. Efficient algorithms for learning Bayesian network structure and parameters from training data have been a focus of much current research. These algorithms generate a single Bayesian model by maximizing the posterior probability or likelihood. In realistic problems, due to the sparse amount of observed data compared to the size of the network, it is often biased to select one model and ignore the model uncertainty. Thus, it is preferable to adopt a full Bayesian approach to account for model uncertainty.

Besides training data, the prior background knowledge provides many ways to adjust uncertainties. The prior background knowledge includes qualitative and quantitative knowledge which describes the entities and their relationships with different levels of abstraction. Quantitative knowledge can be exemplified by a probability elicitation procedure by a domain expert. In most domains, this is particular difficult due to the limitations of expert knowledge in this level. In contrast, qualitative knowledge, which only provides loose constraints with uncertainty on the entities and their relations exist in many science and engineering domains. As recently shown, it can be applied to the learning process [1]. For example, in biomedicine, the statement: "*Gene CTGF, IL11 and OPN cooperatively activate bone metastasis in breast cancer*", entities are gene *CTGF, IL11, OPN* and *Bone metastasis in breast cancer*, their qualitative relation: *cooperatively activate*. In some cases, there are properties which further specify the qualitative relationship. In "*The risk of lung cancer among smokers is approximate 10 times higher than non-smokers*", *Smoking cause lung cancer* and the influence is *10 times higher to non-smokers*. Model uncertainty represented by the qualitative knowledge enables the full Bayesian approach with model averaging. However, one significant drawback of qualitative knowledge is its potential inconsistency. In the same domain, there may exist contradicting qualitative statements on dependency, causality and parameters over a set of entities. Therefore, methods for integrating inconsistent qualitative knowledge and making use of it as prior background knowledge in modeling Bayesian networks and performing quantitative prediction are definite beneficial to the Bayesian framework. Several qualitative reasoning algorithms have been proposed to perform qualitative inference in a Bayesian net-

work [2, 3]. These algorithms perform qualitative inference with sign propagation. In this paper, we propose a novel framework for performing quantitative Bayesian inference with model averaging based on inconsistent qualitative statements. Our method interprets the qualitative statements by a vector of knowledge features whose structure can be represented by a hierarchical Bayesian network. The prior probability for each qualitative knowledge component is calculated based on the hierarchical knowledge model. These knowledge components define Bayesian model classes in the hyperspace. Within each class, a set of constraints on the ground Bayesian model space can be generated. Therefore, the distribution of the ground model space can be decomposed into a set of weighted distributions determined by each model class. This framework is used to perform full Bayesian inference which can be approximated by Monte Carlo methods, but is analytically tractable for smaller networks and statement sets.

In section 2, we propose the hierarchical knowledge model for modeling and integrating qualitative knowledge. In section 3, we describe the quantitative Bayesian inference method with model averaging. In section 4, we apply our method to the ASIA network and perform inference based on a set of inconsistent qualitative statements. Conclusions and further discussion are provided in section 5.

2. Hierarchical inconsistent qualitative knowledge integration

In this section, we introduce a set of knowledge features for translating qualitative statement into a distribution constraint on the Bayesian model space. Then, we propose a hierarchical Bayesian model to represent the qualitative features. Thirdly, we show that the prior probability of a knowledge component can be calculated as a product of the conditional probabilities of these dependent knowledge features.

2.1. Qualitative knowledge feature

The body of qualitative knowledge can be represented by a set of knowledge features which define the structural and parametric constraints on the hyperspace of a set of ground models.

The structure of a graphical network consisting node B and node A can be described by structural qualitative knowledge features with two first-order logic predicates:

$$Depend(A, B) = 0/1 \qquad Influence(A, B) = 0/1 \tag{1}$$

which describe whether A and B are dependent and whether the influence direction is from A to B; *Depend* and *Influence* are denoted by Dp and I , as well as, the set of structural knowledge features is denoted by $\Pi = \{Dp, I\}$.

2.1.2. Parameter qualitative knowledge features

Under each structural feature, the structure dependent parameter knowledge Λ, then can be described by two dependent set of features, i.e. baseline qualitative knowledge features and extended qualitative knowledge features.

Baseline qualitative knowledge feature

Baseline qualitative knowledge features, Σ, define the basic properties of qualitative causal influences and their synergy.

1. Single Influence

Definition 2.1. If a child node B has a parent node A and the parent imposes an isolated influence on the child, then qualitative influence between parent and child is referred to as *Single Influence*. Single influence can be further classified into single positive influence and single negative influence.

Definition 2.2. If presence of parent node A renders presence of child node B more likely, then the parent node is said to have a *Single Positive (SP)* influence on the child node. This can be represented by the inequality

$$Pr(B|A) \geq Pr(B|\overline{A}) \tag{2}$$

Definition 2.3. If presence of parent node A renders presence of child node B less likely, then parent node is said to have a *Single Negative (SN)* influence on child node. This can be represented by the inequality

$$Pr(B|A) \leq Pr(B|\overline{A}) \tag{3}$$

2. Joint Influence

Definition 2.4. If a child node B has more than one parent node and all parents affect the child in a joint way, then these influences between parents and child are referred to as joint influence. This joint influence can be either synergic (cooperative)

or antagonistic (competitive) and the individual influences from the parent to the child can be either positive or negative.

Definition 2.5. If a joint influence from two or more parent nodes generates a combined influential effect larger than the single effect from each individual parent, then the joint influence is referred to as *Plain Synergy (PlSyn)*.

Assume that parent nodes A and B impose positive individual influences on child node C, then the knowledge model can be defined as

$$Pr(C|A,B) \geq \left\{ \begin{array}{c} Pr(C|A,\overline{B}) \\ Pr(C|\overline{A},B) \end{array} \right\} \geq Pr(C|\overline{A},\overline{B}) \tag{4}$$

Definition 2.6. If joint influences from two or more parent nodes generate an combined influential effect larger than the sum of all effects from an individual parent, then the joint influence is referred to as *Additive Synergy (AdSyn)*.

Assume in case that parent nodes A and B impose a positive individual influence on child node C, then we define

$$\begin{aligned} Pr(C|A,B) &\geq Pr(C|A,\overline{B}) + Pr(C|\overline{A},B) \\ &\geq \left\{ \begin{array}{c} Pr(C|A,\overline{B}) \\ Pr(C|\overline{A},B) \end{array} \right\} \geq Pr(C|\overline{A},\overline{B}) \end{aligned} \tag{5}$$

Similar rules can be applied to the case where A and B impose a negative individual influence on child node C.

Comparing Eq. 5 with Eq. 4, we can conclude that *additive synergy* is a sufficient condition for *plain synergy* and *plain synergy* is a necessary but not sufficient condition for *additive synergy*. Therefore, if multiple parents demonstrate additive synergy, it is sufficient to judge that this influence is also plain synergy, but not vice-versa.

It is important to distinguish between plain synergy and additive synergy since they represent distinct semantic scenarios in a domain. For example, A is a protein and B is a kinase which phosphorylates protein A and produces the phosphorylated protein C. Because of the nature of this protein-protein interaction, neither B nor A alone can significantly increase the presence of C, but both together can drastically increase the presence of C which is greater than the sum of C in case of either A or B present. In this example A and B exhibit additive synergy and it is sufficiently to conclude that A and B has plain synergy as well.

Definition 2.7. If the joint influences from two or more parent nodes generate a combined influential effect less than the single effect from individual parent, then the joint influence is referred to as *antagonistic joint influence* or *antagonism (Ant)*.

Assume that parent nodes A and B have independent and symmetric positive single influences on child node C, the antagonistic influence of A and B can be represented by

$$\left\{ \begin{array}{c} Pr(C|A,\overline{B}) \\ Pr(C|\overline{A},B) \end{array} \right\} \geq Pr(C|A,B) \geq Pr(C|\overline{A},\overline{B}) \tag{6}$$

Similar rules can be applied to the case where A and B impose a negative individual influence on child node C.

3. Mixed Joint Influence In case that the joint effect on a child is formed by a mixture of positive and negative individual influences from its parents, the extraction of a probability distribution is not well-defined in general. Hence, if no additional information is given, mixed influences are treated as independent and with equal influential strength. Assume that parent node A imposes positive single influence on child node C and parent node B imposes negative single influence on child node C, then the joint influence or *mixed synergy (MxSyn)* can be represented by

$$\begin{array}{cc} Pr(C|A,B) \geq Pr(C|\overline{A},B) & Pr(C|A,\overline{B}) \geq Pr(C|\overline{A},\overline{B}) \\ Pr(C|A,\overline{B}) \geq Pr(C|A,B) & Pr(C|\overline{A},\overline{B}) \geq Pr(C|\overline{A},B) \end{array} \tag{7}$$

Any additional structure can be brought into the CPT of the corresponding collider structure as soon as dependencies between influences are made explicit by further qualitative statements.

Extended qualitative knowledge feature

The extended qualitative knowledge features, denoted by Ψ, include the relative ratio and difference between any probability configurations of a influence and the absolute bound of any single configuration in a influence. These extended features impose further constraints based on the baseline features and therefore restrain the uncertainty in Bayesian model space more accurately. The extended qualitative knowledge features can be consistently represented by a linear inequality. In the case that node A imposes a single influence on node B, the linear constraints can then be written as

$$Pr(B|A) \geq R \times Pr(B|\overline{A}) + \Delta \tag{8}$$

$Pr(B|A), Pr(B|\overline{A}) \in [Bd_{min}, Bd_{max}]$. R is the *Influence Ratio*, Δ is the *Influence Difference* and Bd represent the *bounds* of the probabilities. Once the qualitative knowledge is translated by the feature set $\{\Pi(Dp,I), \Lambda(\Sigma, \Psi(R, \Delta, Bd))\}$ according to Eq. 1 to Eq. 8, the distribution of ground models is defined by this knowledge.
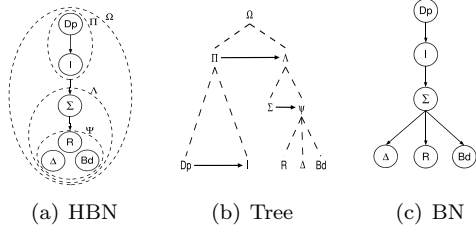
| | (a) HBN | | (b) Tree | | (c) BN | |

Fig. 1: Hierarchical Bayesian Network on Qualitative Knowledge.

Stat.	Dp	I	Σ	R	Δ	B	Weight
S_1	1	1	SP	$[10,\infty]$	null	null	w_1
S_2	1	1	SP	$[25,\infty]$	null	null	w_2
S_3	1	0	null	null	null	null	w_3

Fig. 2: Feature-vector of Statements.

2.2. Hierarchical knowledge model

The dependent qualitative knowledge feature set can be represented by a hierarchical Bayesian network (HBN) [4]. Within a knowledge HBN, the structureal feature Π and parameter feature Λ are two first-level composite nodes. Π can be futher decomposed into two leaf nodes Dp and I. The parameter feature Λ contains two second-level composite nodes, i.e. the baseline knowledge features Σ and extended knowledge features Ψ which consists of three leaf nodes R, Δ and Bd. Thus qualitative knowledge Ω can be described as $\Omega = \{\Pi(Dp, I), \Lambda(\Sigma, \Psi(R, \Delta, Bd))\}$, where $\Sigma = (SP, SN, PlSyn, AdSyn, Ant, MxSyn)$. The hierarchical knowledge model is shown in Figure 1(a) and a tree hierarchy in Figure 1(b). The equivalent Bayesian network is shown in Figure 1(c).

Hierarchical Bayesian Networks encode conditional probability dependencies in the same way as standard Bayesian Networks. The prior probability of a qualitative knowledge Ω can be written as a joint probability of $\{\Pi, \Lambda\}$ and can be decomposed according to the dependency between each component features as follows.

$$Pr(\Omega) = Pr(\Pi)Pr(\Sigma|\Pi)Pr(\Psi|\Sigma) \qquad (9)$$

where $Pr(\Psi|\Sigma) = Pr(R|\Sigma)Pr(\Delta|\Sigma)Pr(Bd|\Sigma)$, $Pr(\Pi) = Pr(Dp)Pr(I|Dp)$ and $Pr(\Sigma|\Pi) = Pr(\Sigma|I)$. The conditional probabilities of qualitative knowledge features can be calculated by counting the weighted occurrences given a set of inconsistent statements. The weight of knowledge features equals to the credibility of their knowledge sources which may be evaluated by a domain ex-

pert or determined by the source *impact factor*. If no further information on the weights is available, they are set to 1. In this case, the conditional probability of features is computed only by occurrence count. For example, we assume a set of qualitative statements, $\widetilde{S} = \{S_1, S_2, S_3\}$, about *smoking* and *lung cancer* are observed: 1) *The risk is more than 10 times greater for smokers to get lung cancer than no-smokers.* 2) *Men who smoke two packs a day increase their risk more than 25 times compared with non-smokers.* 3) *There is not significant evidence to prove that smoking directly cause lung cancer, however, clinical data suggest that lung cancer is related to smoking.* The statements can be represented by a vector of features which is shown in Figure 2. The conditional probability of the features can be calculated straightforwardly by $Pr(I|Dp)=(w_1+w_2)/w_a$, $Pr(\bar{I}|Dp)=w_3/w_a$, $Pr(r_1|SP)=w_1/w_b$ and $Pr(r_2|SP)=(w_1+w_2)/w_b$ where $w_a = w_1 + w_2 + w_3$, $w_b = 2w_1 + w_2$, $Pr(Dp) = 1$, $Pr(SP|I) = 1$, $r_1 = [10, 25]$ and $r_2 = [25, \infty]$. One notion is that the knowledge features $\Psi = \{R, \Delta, Bd\}$ in Figure 1(a) are continuous-valued and therefore, can be transformed to discrete attributes by dynamically defining new discrete attributes that partition the continuous feature value into a discrete set of intervals. In the above example, the continuous feature R in S_1 has value range $[10, \infty]$ and a continuous value range $[25, \infty]$ in S_2. The continuous ranges can be partitioned into two discrete intervals: $r_1 = [10, 25]$ and $r_2 = [25, \infty]$, therefore, the qualitative knowledge $\widetilde{\Omega} = \{\Omega_1, \Omega_2, \Omega_3\}$ can be transformed from $\widetilde{S} = \{S_1, S_2, S_3\}$ with discrete-valued features.

2.3. Qualitative knowledge integration

Once we have calculated the conditional probabilities of knowledge features, the prior probability of qualitative knowledge can be computed according to Eq. 9. Thus the inconsistent knowledge components are ready to be reconciled. The qualitative knowledge transformed from the feature vector of statements in Figure 2 can be described by $\widetilde{\Omega}$: $\Omega_1=\{1, 1, SP, [10, 25], \emptyset, \emptyset\}$, $\Omega_2=\{1, 1, SP, [25, \infty], \emptyset, \emptyset\}$ and $\Omega_3=\{1, 0, \emptyset, \emptyset, \emptyset, \emptyset\}$, where $\Omega_k=\{Dp_k, I_k, \Sigma_k, R_k, \Delta_k, Bd_k\}$. If the weights of statements are set to 1, the knowledge prior probability is calculated, then we have $Pr(\Omega_1)=2/9$, $Pr(\Omega_2)=4/9$ and $Pr(\Omega_3)=1/3$. The integrated qualitative knowledge thus preserved the uncertainty from each knowledge component.

Each qualitative knowledge component Ω_k defines a model class with a set of constraints on the ground model space which is generated by its features. The model class and its constraints are used for modeling Bayesian networks and performing quantitative inference.

3. Bayesian inference with inconsistent qualitative knowledge

In this section, we propose a novel approach to make use of a set of inconsistent qualitative statements and their prior belief distribution as background knowledge for Bayesian modeling and quantitative inference.

A Bayesian model m represents the joint probability distribution of a set of variables $X = \{x_1, x_2, ..., x_N\}$ [5]. The model is defined by a graph structure s and a parameter vector θ, i.e. $m = \{s, \theta\}$. In full Bayesian framework, all available information is used in an optimal way to perform inference by taking model uncertainty into account. Let us classify the set of available information into an available set of training data D and a set of inconsistent qualitative background knowledge $\widetilde{\Omega} = \{\Omega_1, \ldots, \Omega_K\}$ on a constant set of variables. The posterior distribution of models m is then given by

$$Pr(m|D, \widetilde{\Omega}) = \frac{Pr(D|m, \widetilde{\Omega})Pr(m|\widetilde{\Omega})Pr(\widetilde{\Omega})}{Pr(D, \widetilde{\Omega})} \quad (10)$$

The first term in the numerator of Eq. 10 is the likelihood of the data given the model. The second term denotes the model prior which reflects the inconsistent set of background knowledge and the last term is the prior belief of the knowledge set. Now, inference in the presence of evidence is performed by building the expectation across models:

$$Pr(X|D, E, \widetilde{\Omega})$$
$$= \int dm Pr(X|E, m)Pr(D|m, \widetilde{\Omega})Pr(m|\widetilde{\Omega})Pr(\widetilde{\Omega}) \quad (11)$$

In this paper we consider the extreme case of no available quantitative data, $D = \emptyset$.

$$Pr(X|E, \widetilde{\Omega}) = \int dm Pr(X|E, m)Pr(m|\widetilde{\Omega})Pr(\widetilde{\Omega}) \quad (12)$$

In this case, model prior distribution $Pr(m|\widetilde{\Omega})$ is determined soly by the inconsistent background

knowledge set $\widetilde{\Omega}$. Each independent qualitative knowledge component, $\Omega_k \in \widetilde{\Omega}$, uniquely defines a model class, M_k, with a vector of features, i.e. $\widetilde{M} = \{M_1, \ldots, M_K\}$. The features are translated into a set of constraints which determine the distribution of the ground models within each model class.

First of all, the probability of a model class given the inconsistent knowledge set is written as

$$Pr(M_k|\widetilde{\Omega}) = \sum_{i=1}^{K} Pr(M_k|\Omega_i)Pr(\Omega_i|\widetilde{\Omega}) = Pr(\Omega_k) \quad (13)$$

where $\{Pr(M_k|\Omega_i) = 1, i = k\}$ and $\{Pr(M_k|\Omega_i) = 0, i \neq k\}$ since the k-th model class is uniquely defined by Ω_k and is independent to the other knowledge component. Secondly, the probability of a ground Bayesian model sample m in the k-th model class given the inconsistent knowledge set is

$$Pr(m \in M_k|\widetilde{\Omega}) = Pr(m|M_k)Pr(M_k|\widetilde{\Omega}) \quad (14)$$

Thus, the inference on X given evidence E and inconsistent knowledge set $\widetilde{\Omega}$ in Eq. 12 can be written as

$$Pr(X|E, \widetilde{\Omega}) = \sum_k \int_m dm Pr(X|m, E)Pr(m|M_k)Pr(\Omega_k) \quad (15)$$

where $Pr(m|\widetilde{\Omega}) = \sum_k Pr(m \in M_k|\widetilde{\Omega})$ and we assume the inconsistent knowledge set to be true, i.e. $Pr(\widetilde{\Omega}) = 1$. Therefore, the inference is calculated by firstly integrating over the structure space and the structure-dependent parameter space of a ground Bayesian model from a model class according to the constraints and perform such integration iteratively over all possible model classes with the prior distribution. The integration in Eq. 15 is non-trivial to compute, however, Monte Carlo methods can be used to approximate the inference.

4. Empirical study

In this section, we apply our framework to integrate a set of inconsistent qualitative knowledge about ASIA network and perform quantitative Bayesian inference.

4.1. Inconsistent knowledge integration

The ASIA network [6] is a popular toy belief model for testing Bayesian algorithms. The structure and

10. *Significantly more patients with CLD than LC experienced breathlessness in the final year (94% CLD vs 78% LC, P < 0.001) and final week (91% CLD vs 69% LC, P < 0.001) of life.* [8]
11. *95% of patients with chronic bronchitis and emphysema reported Dyspnea.* [9]

Node Pair	CPT β			
(Pa,Ch)	$Pr(Ch|Pa)$		$Pr(Ch|\overline{Pa})$	
(V,T)	$\alpha_1=0.05$		$\alpha_0=0.01$	
(S,LC)	$\beta_1=0.3$		$\beta_0=0.01$	
(S,B)	$\gamma_1=0.6$		$\gamma_0=0.3$	
$(T/LC,X)$	$\lambda_1=0.9$		$\lambda_0=0.05$	
$([Pa_1,Pa_2],Ch)$	$Pr(Ch|Pa_1,Pa_2)$	$Pr(Ch|\overline{Pa_1},Pa_2)$	$Pr(Ch|Pa_1,\overline{Pa_2})$	$Pr(Ch|\overline{Pa_1},\overline{Pa_2})$
$([T/LC,B],D)$	$\zeta_1=0.9$	$\zeta_2=0.7$	$\zeta_3=0.8$	$\zeta_4=0.1$
$([T,LC],T/LC)$	$f_1=1$	$f_2=1$	$f_3=1$	$f_4=0$

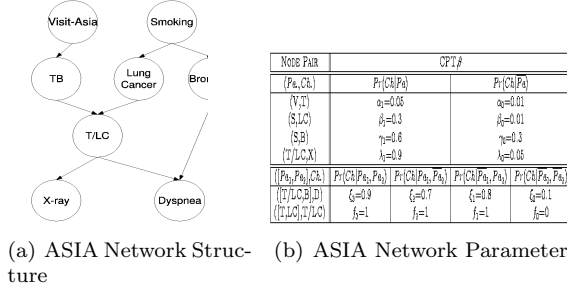(a) ASIA Network Structure   (b) ASIA Network Parameter

Fig. 3: ASIA Belief Network.

parameter of actual ASIA network is shown in Figure 3. For demonstration, we consider the inconsistent qualitative statements with regarding to single edge between *Smoking* and *Lung Cancer*, as well as the collider structure of *Lung Cancer*, *Bronchitis* and *Dyspnea*. The method applies to all of the entities and their relations in the ASIA network.

1. *Although nonsmokers can get lung cancer, the risk is about 10 times greater for smokers.* (www.netdoctor.co.uk)
2. *The lifetime risk of developing lung cancer in smokers is approximately 10%.* (www.chestx-ray.com/Smoke/Smoke.html)
3. *Men who smoke two packs a day increase their risk more than 25 times compared with non-smokers.* (www.quit-smoking-stop.com/lung-cancer.html)
4. *Lifetime smoker has a lung cancer risk 20 to 30 times that of a nonsmoker.* (www.cdc.gov/genomics/hugenet/ejournal/O GGSmoke.htm)
5. *15% of smokers ultimately develop lung cancer.* (www.cdc.gov/genomics/hugenet/ejournal/O GGSmoke.htm)
6. *The mechanisms of cancer are not known. It is NOT possible to attribute a cause to effects whose mechanisms are not fully understood.* (www.forces.org/evidence/evid/lung.htm)
7. *It is estimated that 60% of lung cancer patients have some dyspnea at the time of diagnosis rising to 90% prior to death.* (www.lungcancer.org/health_care/focus_on_ic /symptom/dyspnea.htm)
8. *Muers et al. noted that breathlessness was a complaint at presentation in 60% of 289 patients with non-small-cell lung cancer. Just prior to death nearly 90% of these patients experienced dyspnea.* [7]
9. *At least 60% of stage 4 lung cancer victims report dyspnea.* (www.lungdiseasefocus.com/lung-cancer/

Each statement is analyzed by the hierarchical knowledge model in Figure 1(a) and the extracted features are summarized in Figure 4(a). In this statement set, the first six statements represent the relation between (tobacco)smoking and lung cancer. $\{S_1,\ldots,S_5\}$ describe a *single positive (SP)* influence from smoking to lung cancer with inconsistent knowledge features of the *ratio (R)* and *bound (Bd)*. However, statement S_6 declares a contradicting knowledge suggesting that smoking is not the cause of lung cancer. $\{S_7,\ldots,S_{11}\}$ describe the synergic influence from lung cancer and bronchitis to dyspnea. Without further information, it can be represented by *plain synergy with positive individual influence*. The knowledge on the extended features in Eq. 7 of the conditional probability distribution of this collider structure is not available, however, the knowledge on the extended features of the marginalized conditinal probability space are provided in these statements. For simplicity, we assume the weight of every qualitative statement equals to 1, i.e. $\{w_i = 1, i = 1,\ldots,11\}$. Due to the parameter independency [5], we can compute the conditional probability of each local structure independently. For each local structure, we calculate the conditional probability of knowledge features by counting its occurrence frequency. For the local structure of smoking and lung cancer in the ASIA network, the prior probability of the knowledge features can be calculated as $Pr(Dp)=5/6$, $Pr(I|Dp)=1$, $Pr(\overline{I}|\overline{Dp})=1$, $Pr(SP|I)=1$, $Pr(r_1|SP)=1/5$, $Pr(r_2|SP)=1/5$, $Pr(r_3|SP)=2/5$, $Pr(r_4|SP)=1/5$, $Pr(b_1|SP)=1/2$ and $Pr(b_2|SP)=1/2$. where $r_1 = [9,11]$, $r_2 = [20,25]$, $r_3 = [25,30]$ and $r_4 = [30,\infty]$; $b_1 = [9\%,11\%]$ and $b_2 = [14\%,16\%]$. The continuous-valued feature R and Bd are discretized into $|R| = 4$ and $|Bd| = 2$ discrete-value intervals respectively. Based on the features and their prior belief, a set of qualitative knowledge $\widehat{\Omega} = \{\Omega_1,\ldots,\Omega_{16}\}$ is formed in Figure 4(b).

Stat.	Dp	I	Σ	R	Δ	Bd	Weight
S_1	1	1	SP	[9, 11]	null	null	w_1
S_2	1	1	SP	null	null	[9%, 11%]	w_2
S_3	1	1	SP	[25, ∞]	null	null	w_3
S_4	1	1	SP	[20, 30]	null	null	w_4
S_5	1	1	SP	null	null	[14%, 16%]	w_5
S_6	0	0	SP	null	null	null	w_6
S_7	1	1	SP	null	null	[60%, 90%]	w_7
S_8	1	1	SP	null	null	[60%, 90%]	w_8
S_9	1	1	SP	null	null	[60%, 100%]	w_9
$S_{10(1)}$	1	1	SP	null	null	[69%, 78%]	w_{10}
$S_{10(2)}$	1	1	SP	null	null	[91%, 94%]	w_{10}
S_{11}	1	1	SP	null	null	[94%, 96%]	w_{11}

(a) Feature-vector of Statements

Ω	Dp	I	Σ	R	Δ	Bd	$Pr(\Omega_i)$
Ω_1	1	1	SP	[9, 11]	null	[9%, 11%]	1/12
Ω_2	1	1	SP	[9, 11]	null	[14%, 16%]	1/12
Ω_3	1	1	SP	[20, 25]	null	[9%, 11%]	1/12
Ω_4	1	1	SP	[20, 25]	null	[14%, 16%]	1/12
Ω_5	1	1	SP	[25, 30]	null	[9%, 11%]	1/6
Ω_6	1	1	SP	[25, 30]	null	[14%, 16%]	1/6
Ω_7	1	1	SP	[30, ∞]	null	[9%, 11%]	1/12
Ω_8	1	1	SP	[30, ∞]	null	[14%, 16%]	1/12
Ω_9	0	0	null	null	null	null	1/6
Ω_{10}	1	1	PISyn	null	null	null	1
Ω_{11}	1	1	SP	null	null	[60%, 69%]	3/11
Ω_{12}	1	1	SP	null	null	[69%, 78%]	4/11
Ω_{13}	1	1	SP	null	null	[78%, 90%]	3/11
Ω_{14}	1	1	SP	null	null	[90%, 100%]	1/11
Ω_{15}	1	1	SP	null	null	[91%, 94%]	1/2
Ω_{16}	1	1	SP	null	null	[94%, 96%]	1/2

(b) Integrated Qualitative Knowledge with Prior Probability

Fig. 4: Qualitative Statements and Knowledge in ASIA network.

4.2. ASIA model monte carlo sampling

Given the integrated qualitative knowledge set $\widetilde{\Omega}$ with prior probabilities, we now construct the Bayesian model class and the distribution on ground model space within each class. For demonstration purposes, we assume the partial structure and its parameters, i.e. $\{\alpha, \gamma, \lambda, f\}$, to be known as in Figure 3(b). Therefore the uncertainty of ASIA model space is restricted to the uncertainty of the local structure and parameter space on *Smoking* and *Lung Cancer* which can be described by $Pr(m|M_k)$ and $Pr(M_k)$ defined by $\{\Omega_k | k = 1, \ldots, 9\}$, i.e. $\{M_k(\Omega_k)|k = 1, \ldots, 9\}$, as well as the uncertainty of the local space on *Lung Cancer*, *Bronchitis* and *Dyspnea* which can be jointly determined by three types of model class, i.e. the root-dimension model class defined by Ω_{10}, the marginal-dimension model classes of lung cancer and dyspnea defined by $\{\Omega_i | i = 11, \ldots, 14\}$ and the marginal-dimension model classes of bronchitis and dyspnea defined by $\{\Omega_j | j = 15, 16\}$. Thus, there are total eight possible combination of these model classes, i.e. $\{M_k(\Omega_{10}, \Omega_i, \Omega_j)|k = 10, \ldots, 17; i = 11, \ldots, 14; j = 15, 16\}$ and each combination virtually forms a complete model class which defines the set of constraints on the structure and parameter space of ground Bayesian model for the local collider structure of lung cancer, bronchitis and dyspnea. The prior probability of each combination, $Pr(M_k)$ is the product of the prior probability of its independent components, i.e.

$$Pr(M_k) = Pr(\Omega_{10})Pr(\Omega_i)Pr(\Omega_j) \qquad (16)$$

For each local structure, we perform 10,000 sampling iterations. In each iteration, we select a model class M_k randomly based on the prior probability of the model class, i.e $Pr(M_k)$. In each selected model class, we randomly choose 3 samples of ground Bayesian model m, whose structure and parameter space is consistent with the class constraints $Pr(m|M_k)$ as shown in Figure 1(a). In this way, for the local structure of smoking and lung cancer, the prior bability of the model class is equivalent to its knowledge component, i.e. $Pr(M_k)=Pr(\Omega_k)$. We generate total N=30,000 ground model samples from model classes $\{M_k(\Omega_k)|k = 1, \ldots, 9\}$ defined by Ω_k in Figure 4(b). The ground model samples are shown in Figure 5(a). For the local collider structure of lung cancer, bronchitis and dyspnea, we generate N=30,000 ground model samples from the combination of model classes defined in Eq. 16 based on $\{\Omega_k|k = 10, \ldots, 16\}$ in Figure 4(b). The marginal conditional probability samples are shown in Figure 5(b) and 5(c). Without further information on lung cancer, bronchitis and dyspnea, we can set their prior probabilities to be 1/2. By taking average over the models in Figure 5(a) to 5(c), we can calculate the mean value for the conditional probability of lung cancer given smoking, i.e. $\overline{\beta_1}$=0.1255, $\overline{\beta_0}$=0.006, and of Dyspnea given lung cancer and Bronchitis, i.e. $\overline{\xi_0}$=0.2725, $\overline{\xi_1}$=0.9053, $\overline{\xi_2}$=0.5495 and $\overline{\xi_3}$=0.968. Note that since the *9th* model class defined by Ω_9 for the structure of lung cancer and smoking, i.e. $M_9(\Omega_9)$, contains no edge between the nodes, the parameter of this model class is null.
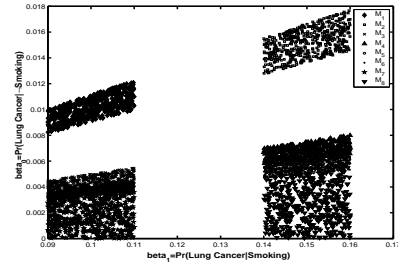
4.3. ASIA model inference

For each of the model sample, according to Eq. 15, we perform inferences *in silico* on the likelihood of a patient having lung cancer (Lc) given information about the patient's smoking status and clinical evidences including observation of X-ray, Dyspnea, and Bronchitis, i.e. $X_{obs} = \{Sm, Xr, Dy, Br\}$. The convergence of these prediction under a set of evidences $\widetilde{E} = \{E_1, E_2, E_3, E_4, E_5, E_6\}$ are shown in Figure 5(d). The true prediction values with parameters in Figure 3(b) under the evidence set $\widetilde{E}$ are listed below in Figure 6. The presence of bronchitis could explain away the probability of lung cancer and the presence of smoking increases the risk of getting lung cancer.

5. Conclusions

In this paper, we proposed a hierarchical Bayesian model for representing qualitative knowledge with a vector of features. The inconsistent knowledge components are integrated by calculating a prior distribution. The integrated qualitative knowledge set is used as prior background knowledge in modeling Bayesian networks and performing quantitative inference. Simulation results suggest that our methods can produce reasonable quantitative prediction based on the inconsistent knowledge set.

References

[1] R. Chang and M. Stetter. Quantitative bayesian inference by qualitative knowledge modeling. In *Proceedings of the 20th International Joint Conference on Neural Networks*, 2007.

[2] M. Henrion M.J. Druzdzel. Efficient reasoning in qualitative probabilistic networks. In *Proceedings of the Eleventh National Conference on Artificial Intelligence*, pages 548–553, Washington DC, 1993. AAAI Press.

[3] S. Parsons S. Renooij, L.C. van der Gaag. Context-specific sign-propagation in qualitative probabilistic networks. *Artificial Intelligence*, 140:207–230, 2002.

[4] P. Flach E. Gyftodimos. Hierarchical bayesian networks: A probabilistic reasoning model for structured domains. In *Proceedings of the ICML-2002 Workshop on Development of Representations*, pages 23–30, 2002.

[5] D. Heckerman. A tutorial on learning with bayesian networks. Technical report, Microsoft Research, USA, 1996.

[6] D. J. Spiegelhalter S. L. Lauritzen. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistics Society B*, 50(2):157–194, 1988.

[7] M. Lertzman D. J. Dudgeon. Dyspnea in the advanced cancer patient. *J. of Pain and Symptom Management*, 16(4):212–219, 1998.

[8] S. Khan P. Edmonds, S. Karlsen and J. Addington-Hall. A comparison of the palliative care needs of patients dying from chronic respiratory diseases and lung cancer. *Palliative Medicine*, 15(4):287–295, 2001.

[9] E. Fernandez J.F. Dirks M. Schocket R.A. Kinsman, R.A. Yaroush and J. Fukuhara. Symptoms and experiences in chronic bronchitis and emphysema. *Chest*, 83:755–761, 1983.
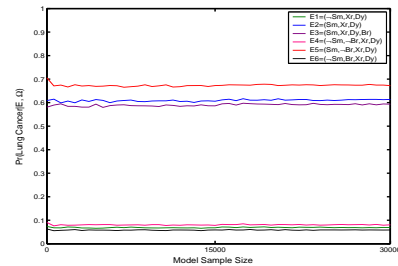
(a) Model Samples for Smoking and Lung Cancer
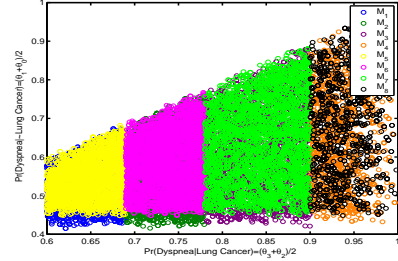
(b) Model Samples for Lung Cancer and Dyspnea
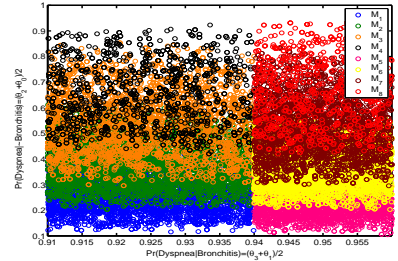
(c) Model Samples for Bronchitis and Dyspnea

(d) Quantitative Bayesian Inference

Fig. 5: ASIA Model Sampling and Inference.

Exp.	E_1	E_2	E_3	E_4	E_5	E_6
True	0.17	0.87	0.84	0.21	0.91	0.11
Simulation	0.07	0.61	0.59	0.08	0.67	0.06

Fig. 6: Inference Results on ASIA Network