

Inducing Probability Distributions from Knowledge Bases with (In)dependence Relations

Jianbing Ma¹ and Weiru Liu¹ and Anthony Hunter²

¹School of Electronics, Electrical Engineering and Computer Science,
Queen's University Belfast, Belfast BT7 1NN, UK
{jma03, w.liu}@qub.ac.uk

²Department of Computer Science, University College London,
Gower Street, London WC1E 6BT, UK
a.hunter@cs.ucl.ac.uk

Abstract

When merging belief sets from different agents, the result is normally a consistent belief set in which the inconsistency between the original sources is not represented. As probability theory is widely used to represent uncertainty, an interesting question therefore is whether it is possible to induce a probability distribution when merging belief sets. To this end, we first propose two approaches to inducing a probability distribution on a set of possible worlds, by extending the *principle of indifference on possible worlds*. We then study how the (in)dependence relations between atoms can influence the probability distribution. We also propose a set of properties to regulate the merging of belief sets when a probability distribution is output. Furthermore, our merging operators satisfy the well known Koniczny and Pino-Pérez postulates if we use the set of possible worlds which have the maximal induced probability values. Our study shows that taking an induced probability distribution as a merging result can better reflect uncertainty and inconsistency among the original knowledge bases.

Introduction

In general, operators for merging multiple belief/knowledge bases (flat or stratified) in the literature can be divided into two families: syntax-based (e.g., (Baral et al. 1992; Koniczny 2000; Delgrande, Dubois, and Lang 2006), etc) and model-based (e.g., (Koniczny and Pino-Pérez 1998; Benferhat, Lagrue, and Rossit 2007; Qi, Liu, and Bell 2006), etc). For either family, the result of merging is effectively a set of formulae that is normally required to be consistent, and hence it is difficult to reflect inconsistency between the sources in the result of merging. To illustrate, let us consider the following example. Assume that we have two knowledge bases (KBs) representing two agents' beliefs, $K_1 = \{p\}$, and $K_2 = \{\neg p\}$. What should the merging result $\Delta(K_1, K_2)$ be (assume that Δ is a merging operator)? Intuitively, we cannot expect $\Delta(K_1, K_2)$ to be either $\{p\}$ or $\{\neg p\}$, since it is biased towards one agent or the other, whilst letting $\Delta(K_1, K_2)$ be either $\{\top\}$ or $\{\perp\}$ does not give a meaningful result (\top or \perp could come from the merging of $p \wedge q$ versus $\neg p \wedge \neg q$ or any two conflicting pieces of beliefs in this way). In this scenario, it is better to represent the merged belief as a probability distribution $P(p) = P(\neg p) = 0.5$,

which accurately reflects that the two hypotheses considered are mutually conflicting and equal w.r.t support from the original knowledge bases.

Probability theory is one of the most used approaches to modeling and reasoning with uncertain knowledge. However, probability theory has not been directly associated with representing a result of merging of classical knowledge bases. Research efforts that have some connections with this line of research were seen in (Bacchus et al. 1996) and (Knight 2002). In (Bacchus et al. 1996), a *random-worlds method* based on the *principle of indifference* was proposed to induce a probability distribution on a set of possible worlds from a given knowledge base K , such that for $w \models K$, $P(w) = \frac{1}{|Mod(K)|}$ (where $Mod(K)$ is the set of models of K). In (Knight 2002), probability theory was used to measure the degree of (in)consistency (of subsets) of a knowledge base. Assume that \mathcal{P} is the set of all probability distributions definable on the set of possible worlds from a language (Paris 1994), then given an inconsistent KB K (or a subset of K), the η -consistency measure obtains the probability distribution P in \mathcal{P} such that $P(\alpha) \geq \eta, \forall \alpha \in K$ and for any $P' \in \mathcal{P}$, $\exists \beta \in K, s.t., P(\beta) > P'(\beta)$. Nevertheless, there are still no concrete probability distributions derived from a given knowledge base.

In this paper, we investigate how a probability distribution can be induced as a result of merging multiple KBs, especially for situations where multiple agents' beliefs are inconsistent. To achieve this, we propose two extensions to the *principle of indifference* on possible worlds in (Bacchus et al. 1996) for inducing a probability distribution from a single knowledge base scenario to multiple KBs scenario, namely, the *principle of indifference on knowledge bases* in a knowledge profile and the *principle of indifference on knowledge profile*. This is our first contribution of the paper.

In classical knowledge base merging, (in)dependence relationships between atoms (of a propositional language) are not considered and cannot be considered. For example, if p and q are two atoms, then whether the fact p is true will influence that q is true is not in anyway reflected in either an original knowledge base or a merged result. However, when we consider using a probability distribution as a result of merging, independence relationships among atoms do play a role as to how a final probability distribution can be obtained. To elaborate this further, let us look at two sets of

knowledge bases in the following example.

Example 1 Assume that three people are discussing a person holding a flag spotted at a distance. The three knowledge bases obtained from them are $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, where m for the person is a male and r for the flag is red. An intuitive conclusion is that the flag is red but we do not know whether the person is a male. An underlying assumption for drawing this conclusion in commonsense reasoning is that the color of a flag and the gender of a person are largely independent. Hence the contradicting opinions on the gender of the person do not affect the consistent opinions about the color of the flag.

Now, let us reword this example as follows. Assume that three people are discussing a person's salary and his profession. The three knowledge bases obtained from them are $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, where m for the person is a professor and r for the salary is over \$80k. What would be the result of merging?

In this example, conclusion $\{r\}$ seems largely acceptable for the first scenario because we usually disassociate the attribute *color of a flag* from the attribute *gender of a person*. On the other hand, conclusion $\{r\}$ seems much less acceptable in the 2nd scenario because we know a person's profession and his salary are closely related. A merging operator for classical KBs, such as, $\Delta^{d_D, sum, sum}$ (Konieczny, Lang, and Marquis 2004), $\Delta^{d_H, \Sigma}$ (Revesz 1997), produces the same solution $\{r\}$ for both sets of KBs (scenarios) regardless of if there are any dependencies among atoms (attributes).

Therefore, independence relationships between atoms must be considered, since such independence relationships influence how a probability distribution is defined. To achieve this, we develop two approaches to revising induced probability distributions from the two extensions, with (in)dependence relationships between atoms considered. This is our second main contribution of the paper.

An induced probability distribution from a set of knowledge bases should truly reflect the *beliefs* encoded by these bases. We propose the following three constraints on an induced probability distribution. This is our third contribution.

- A possible world considered possible by some agents should receive a positive probability value, otherwise its probability value is 0. This is the *Positiveness* principle.
- A possible world considered at least as possible as another possible world by all agents shall be assigned with a probability value at least as large as the value assigned to the latter one. This is the *Monotonicity* principle.
- Probability values on possible worlds should be insensitive to syntax. This is the *Syntax-irrelevance* principle.

The rest of the paper is organized as follows. Section 2 introduces preliminaries. In Section 3, we propose two methods for inducing probability distributions. In Section 4, we propose approaches to revising an induced probability distribution considering where some atoms are (in)dependent. In Section 5, we compare our merging framework with logic based merging. Finally, we conclude the paper in Section 6.

Preliminaries

We consider a propositional language $\mathcal{L}_{\mathcal{P}}$ defined from a finite set \mathcal{P} of propositional atoms, denoted by p, q, r etc (possibly with superscripts). An interpretation w (or possible world) is a function that maps \mathcal{P} onto the set $\{0, 1\}$. The set of all interpretations is denoted as W . w is a model of (or satisfies) ϕ iff $w(\phi) = 1$, denoted as $w \models \phi$. We denote $Mod(\phi)$ as the set of models for ϕ . A term on a set of atoms $\{p_1, \dots, p_k\}$ is a formula with the form $p'_1 \wedge \dots \wedge p'_k$ in which literal p'_i is either p_i or $\neg p_i$, $1 \leq i \leq k$.

A (flat) knowledge base K is a finite set of propositions. K is consistent iff there is at least one interpretation that satisfies all propositions in K .

A knowledge profile E is a multi-set of knowledge bases such that $E = \{K_1, K_2, \dots, K_n\}$. $\sqcup E = K_1 \sqcup \dots \sqcup K_n$ denotes the multi-set union of K_i s and $\bigwedge E = K_1 \wedge \dots \wedge K_n$ denotes the conjunction of knowledge bases K_i s of E . E is called consistent iff $\bigwedge E$ is consistent. Let \mathcal{F}_E denote all the atoms appeared in E . $E_1 \leftrightarrow E_2$ denotes that there is a bijection g from $E_1 = \{K_1^1, \dots, K_n^1\}$ to $E_2 = \{K_1^2, \dots, K_n^2\}$ such that $\vdash g(K) \leftrightarrow K$.

Inducing Probability Distributions

In (Bacchus et al. 1996), a random-worlds method was proposed to induce a probability distribution on a set of possible worlds from a single knowledge base. This method is based on the *principle of indifference to possible worlds* which states that given a knowledge base K , all the models of the knowledge base are equally likely. That is, $\forall w \in Mod(K)$, $P(w) = \frac{1}{|Mod(K)|}$ (otherwise $P(w) = 0$) where P is a probability distribution on W . This method cannot be directly applied to induce a probability distribution from a knowledge profile, since the existence of multiple KBs may need to be taken into account. In this section, we propose two extensions to this principle in order to induce a probability distribution from a knowledge profile.

Before proceeding with these extensions, let us first formalize the following three constraints that shall be satisfied by an induced probability distribution, as we discussed in the Introduction. Let E be a knowledge profile and P_E be an induced probability distribution from E .

w-Positiveness: If $\forall K_i \in E, w \not\models K_i$, then $P_E(w) = 0$; else $P_E(w) > 0$.

w-Positiveness states that if a possible world w is considered impossible by all agents, then its induced probability value should be 0; otherwise this value is positive.

w-Monotonicity: For any $K_i \in E$, if $w_1 \models K_i$ implies $w_2 \models K_i$, then $P_E(w_2) \geq P_E(w_1)$. In addition, if $\exists K_j \in E$ such that $w_2 \models K_j$ but $w_1 \not\models K_j$, then $P_E(w_2) > P_E(w_1)$.

w-Monotonicity says that if w_2 is believed at least as plausible as w_1 , then the induced probability value for w_2 should not be less than that assigned to w_1 .

Syntax-Irrelevance: Let α, β be two formulae, then for any w , $P_{E \cup \{\{\alpha \wedge \beta\}\}}(w) = P_{E \cup \{\{\alpha, \beta\}\}}(w)$. If α' is a formula such that $\alpha \equiv \alpha'$, then for any w , $P_{E \cup \{\{\alpha\}\}}(w) = P_{E \cup \{\{\alpha'\}\}}(w)$.

Syntax irrelevance property is a reasonable property in many logic based reasoning systems, such as, knowledge bases merging (Konieczny and Pino-Pérez 1998).

Principle of indifference to knowledge bases

The first extension we attempt to achieve is that for any w , if it is considered possible by a subset of knowledge bases in the profile, then each K in the subset provides an equally likely probability value for w , regardless of which K it is. The final probability value for w is proportional to the sum of such equally likely probability values supporting it. This extension is referred to the *principle of indifference to knowledge bases*. That is, all knowledge bases are treated equally w.r.t. a possible world that is their common model. In another words, each of them contributes the same amount of probability to a possible world that is a model of them.

Let ∇ be an operator mapping knowledge profiles to probability distributions and E be a knowledge profile.

Definition 1 (*Principle of Indifference to Knowledge bases*) Based on this principle, $\nabla(E)$ produces a probability distribution P_E^{IndK} over W s.t., for $w \in W$

$$P_E^{IndK}(w) = \frac{S(w)}{\sum_{w' \in W} S(w')} \quad (1)$$

where $S(w) = |\{K_i \text{ s.t. } K_i \in E, w \models K_i\}|$. $S(w)$ is the number of KBs which consider w possible, and $\sum_{w' \in W} S(w')$ can be rewritten as $\sum_{K_i \in E} |Mod(K_i)|$. Obviously, we have $\sum_{w \in W} P_E^{IndK}(w) = 1$ which shows P_E^{IndK} is a valid probability distribution.

For convenience, hereafter, we omit notation $\nabla(E)$ and simply use a probability distribution as a result of merging.

Proposition 1 P_E^{IndK} satisfies *w-Positiveness*, *w-Monotonicity*, and *Syntax-Irrelevance*.

Example 2 (*Example 1 Cont.*) Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, then from Def. 1, we get $P_E^{IndK}(\neg m \wedge r) = P_E^{IndK}(m \wedge r) = \frac{2}{5}$, $P_E^{IndK}(\neg m \wedge \neg r) = \frac{1}{5}$, $P_E^{IndK}(m \wedge \neg r) = 0$.

In this example, five possible worlds are models of these KBs (i.e., K_1 has two models, K_2 has one and K_3 has two). Possible world $m \wedge r$ has $1/5$ support from K_2 and K_3 respectively. Therefore, it has a probability value $2/5$.

The fact that K_1 has more models than K_2 is not fully considered under this principle. It may be useful to take such a fact into account, especially when a KB K has a large number of models (so K is less specific than other KBs). This leads us to the second extension below.

Principle of indifference to knowledge profile

The second extension we aim to achieve is that for a given K , the contribution from K to w is $\frac{1}{|Mod(K)|}$ if $w \in Mod(K)$ or 0 otherwise. Then for each w , its probability contributions from all the KBs are accumulated, and finally this accumulated value is normalized by the total number of KBs that a profile has. In another word, it is the whole profile (not just the KBs that have contributed to w) determines the final outcome. This extension, is referred to as the *principle of indifference to knowledge profile*.

Definition 2 (*Principle of Indifference to Knowledge Profile*) Based on this principle, a probability distribution P_E^{IndE} over W is defined s.t. for $w \in W$

$$P_E^{IndE}(w) = \frac{1}{|E|} \sum_{i=1}^{|E|} \frac{w(K_i)}{|Mod(K_i)|}. \quad (2)$$

Note that $w(K) = 1$ if $w \models K$ and 0 otherwise (definition of a possible world). Also, Equation 2 requires each K considered is consistent whilst Definition 1 does not require this. In addition, it is easy to check that $\sum_{w \in W} P_E^{IndE}(w) = 1$.

Proposition 2 P_E^{IndE} satisfies *w-Positiveness*, *w-Monotonicity*, and *Syntax-Irrelevance*.

Example 3 (*Example 1 Cont.*) Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, then from Def. 2, we get $P_E^{IndE}(\neg m \wedge r) = \frac{1}{3}$, $P_E^{IndE}(\neg m \wedge \neg r) = \frac{1}{6}$, $P_E^{IndE}(m \wedge r) = \frac{1}{2}$, and $P_E^{IndE}(m \wedge \neg r) = 0$.

With the different points of view captured by the two principles of indifference, the merging result from either Example 2 or Example 3 can be considered acceptable. For instance, with the second extension (Example 3), we have K_1 strongly rejects $m \wedge r$, K_2 strongly supports it, whilst K_3 partially supports it, therefore it is reasonable to obtain $P_E^{IndE}(m \wedge r) = \frac{1}{2}$. Therefore, which approach among the two to use largely depends on the application scenario.

Effect of relationship between atoms

When we induce a probability distribution on a set of possible worlds as the result of merging, we have not yet consider if any atoms defining the language are dependent in terms of probability theory. This is consistent with what has been done in logic-based merging, where relationships between atoms are not considered. However, when we consider probability distributions emerging from a knowledge profile, we must investigate how relationships among atoms will affect an induced probability distribution.

Let us first recall the definition of independence relationship in probability theory.

Definition 3 Let $\{p_1, \dots, p_n\}$ be the set of atoms and P a probability measure on $\{p_1, \dots, p_n\}$. p_1, \dots, p_n are mutually independent w.r.t. P iff for any subset $\{p_{i_1}, \dots, p_{i_k}\}$ of $\{p_1, \dots, p_n\}$, we have $P(p_{i_1} \wedge \dots \wedge p_{i_k}) = \prod_{j=1}^k P(p_{i_j})$.

The necessity of considering (in)dependence relationships: an example

Example 4 (*Example 1 Cont.*) Recall that the conclusion we draw from the first scenario in this example is that the flag is red but we do not know whether the person is a male or not. In terms of probability theory, we expect a probability distribution P that is compatible with this conclusion would generate $P(m) = P(\neg m) = \frac{1}{2}$ (impartial about the person being a male or a female), $P(r) > P(\neg r)$ (the flag is more likely to be red), and $P(m \wedge r) = P(\neg m \wedge r)$ (otherwise we may infer it is a male holding a red flag or a female holding a red flag).

In Example 2, we have $P_E^{IndK}(\neg m \wedge r) = P_E^{IndK}(\neg m \wedge \neg r) = \frac{2}{5}$, $P_E^{IndK}(m \wedge r) = \frac{1}{5}$. Therefore, we get

$P_E^{IndK}(m) = \frac{1}{5}$, $P_E^{IndK}(\neg m) = \frac{4}{5}$, and $P_E^{IndK}(r) = \frac{3}{5}$, $P_E^{IndK}(\neg r) = \frac{2}{5}$. This result is different from our intuitive conclusion and tempts us to believe that the person may not be a male, since $P_E^{IndK}(m) < P_E^{IndK}(\neg m)$.

In Example 3, we have $P_E^{IndE}(\neg m \wedge r) = \frac{1}{3}$, $P_E^{IndE}(\neg m \wedge \neg r) = \frac{1}{6}$, $P_E^{IndE}(m \wedge r) = \frac{1}{2}$. Since $m \wedge r$ has the largest probability value, it may lead us to believe that it is a male holding a red flag which is also inconsistent with our intuition.

The main reason for not being able to draw a satisfactory result from the two induced probability distributions is that the independence assumption between m (a male) and r (a red flag) was not considered. In fact, from the intuitions that $P(m) = P(\neg m) = \frac{1}{2}$, $P(r) > P(\neg r)$, and $P(m \wedge r) = P(\neg m \wedge r)$, we can infer that $P(r|m) = P(r \wedge m)/P(m) = 2P(r \wedge m) = P(r \wedge m) + P(r \wedge \neg m) = P(r)$. That is, r and m are probabilistically independent. This kind of independence relationship among atoms, which cannot be represented in a logic-based representation of knowledge bases, exists in common sense reasoning and real-world applications. Therefore, when considering probability distributions induced from knowledge base merging, such independence relationships must be considered.

All atoms are pair-wise independent

In this subsection, we consider situations where all atoms are pair-wise independent. We start with defining probabilities on atoms, since if all atoms are probabilistically independent, a unique probability distribution on possible worlds (w.r.t Def. 3) can be obtained, and hence probability on formulae can be calculated from their models. It should be pointed out that the following two definitions on induced probability functions on atoms are counterparts of the ones defined in Definitions 1 and 2, respectively. Probability values on atoms derived from a knowledge profile should also satisfy certain constraints. By extending the above constraints on a probability distribution on possible worlds, we have the following revised constraints (Syntax-Irrelevance does not need to be changed).

a-Positiveness If $\forall K_i \in E, K_i \models p$, then $P_E(p) = 1$; if $\forall K_i \in E, K_i \models \neg p$, then $P_E(p) = 0$; else $P_E(p) \in (0, 1)$.

a-Positiveness (atom-Positiveness) states that if all agents believe that p is true, then the merging result should also believe that p is true. Conversely, if all agents think p is false, then the merging result also considers it false. Otherwise p is not totally believed or disbelieved.

a-Monotonicity For any $K_i \in E$, if $K_i \models p_1$ implies $K_i \models p_2$, then $P_E(p_2) \geq P_E(p_1)$. Furthermore, if $\exists K_j \in E$ such that $K_j \models p_2$ but $K_j \not\models p_1$, then $P_E(p_2) > P_E(p_1)$.

a-Monotonicity says that if p_2 is believed at least to the degree of p_1 , then the induced probability value for p_2 should not be less than that assigned to p_1 .

Below, we provide two counterparts of Def. 1 and Def. 2 on atoms.

Definition 4 (Support for p) Let E be a knowledge profile and p be an atom, then the support for p w.r.t E is defined as $Supp_E(p) = \sum_{i=1}^{|E|} |Supp_{K_i}(p)|$ where $Supp_{K_i}(p) = \{w, s.t., w \models p, w \models K_i\}$.

$Supp_{K_i}(p)$ means that we count the number of possible worlds that are models of p and are also models of a knowledge base K_i . That is, p and K_i are consistent since they have at least one common model w , and hence knowledge base K_i can be considered as at least partially supporting p w.r.t w . Then the support for p from the whole profile E is the sum of its support from each K_i .

Based on the definition of support for p , a probability value on p can be calculated as follows.

Definition 5 (Probability value on atom through its support) Let E be a knowledge profile and p be an atom, then a probability value generated from E for p is defined as

$$P_E^{supp}(p) = \frac{|Supp_E(p)|}{|Supp_E(p) + Supp_E(\neg p)|}. \quad (3)$$

Obviously, $Supp_E(p) + Supp_E(\neg p) = \sum_{i=1}^{|E|} |Mod(K_i)|$, and we have $\forall p, P_E^{supp}(p) + P_E^{supp}(\neg p) = 1$ which shows that P_E^{supp} is a valid probability distribution.

Proposition 3 P_E^{supp} satisfies a-Positiveness, a-Monotonicity, and Syntax-Irrelevance.

Example 5 (Example 1 Cont.) Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, then from Def. 5, we get $P_E^{supp}(m) = \frac{2}{5}$, $P_E^{supp}(r) = \frac{4}{5}$.

Observe that $P_E^{supp}(m) = P_E^{IndK}(m)$, $P_E^{supp}(r) = P_E^{IndK}(r)$. This fact holds in general as stated by the following proposition.

Proposition 4 Let E be a knowledge profile, then for any $p \in \mathcal{F}_E$, we have $P_E^{supp}(p) = P_E^{IndK}(p)$, where $P_E^{IndK}(p) = \sum_{w \models p} P_E^{IndK}(w)$.

P_E^{supp} is in fact a counterpart of P_E^{IndK} on atoms where relationships between atoms are considered. However, for a propositional formula μ , generally we do not have $P_E^{supp}(\mu) = P_E^{IndK}(\mu)$ due to the existence of (in)dependence relations.

Definition 6 (Probability value on atom through its ratio of support) Let E be a knowledge profile, and p be an atom, then a probability value generated from E for p is defined as

$$P_E^r(p) = \frac{1}{|E|} \sum_{i=1}^{|E|} \frac{|Supp_{K_i}(p)|}{|Supp_{K_i}(p)| + |Supp_{K_i}(\neg p)|}. \quad (4)$$

It is easy to verify that $\forall p, P_E^r(p) + P_E^r(\neg p) = 1$, hence P_E^r is indeed a valid probability function.

Proposition 5 P_E^r satisfies a-Positiveness, a-Monotonicity, and Syntax-Irrelevance.

Below we show that P_E^r is indeed a counterpart of P_E^{IndE} .

Proposition 6 Let E be a knowledge profile, then for any $p \in \mathcal{F}_E$, we have $P_E^{IndE}(p) = P_E^r(p)$.

Example 6 (Example 1 Cont.) Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{\neg m\}$, $K_2 = \{r \wedge m\}$ and $K_3 = \{r\}$, then from Def. 6, we get $P_E^r(m) = \frac{1}{2}$, $P_E^r(r) = \frac{5}{6}$. Obviously this result is more consistent with our intuition such that $P_E^r(m) = P_E^r(\neg m) = \frac{1}{2}$, $P_E^r(r) > P_E^r(\neg r)$, and $P_E^r(m \wedge r) = P_E^r(\neg m \wedge r)$.

Results from Examples 5 and 6 illustrate that in some scenarios, the principle of indifference to knowledge profile is more reasonable than that to knowledge bases.

Some atoms are independent

Now we investigate situations where some atoms are independent, contrary to what is assumed in the above subsection. This assumption is rational in many real-world applications. Typically when atoms (attributes) are defined, their (in)dependent relationships are clear from context¹.

To proceed, here we assume that the dependence relation between atoms is an equivalence relation which satisfies the following conditions.

- $\forall p, p$ is dependent of p ,
- $\forall p, q, p$ is dependent of q iff q is dependent of p ,
- $\forall p, q, r$, if p is dependent of q and q is dependent of r , then p is dependent of r .

With these conditions, $\mathcal{F}_E = \{p_1, \dots, p_n\}$ can be partitioned by a dependence relation. Let a partition of \mathcal{F}_E be $\mathcal{L} = \{\{p_1^1, \dots, p_{m_1}^0\}, \dots, \{p_1^k, \dots, p_{m_k}^k\}\}$ s.t. for any $p, p' \in \mathcal{F}_E$ if $\exists L \in \mathcal{L}$ where $p, p' \in L$ then p and p' are dependent, otherwise they are independent. Two extremes are either all the atoms are dependent (so there is only one subset in the partition) or all the atoms are independent (then we have exactly n partition groups).

For a subset $L = \{p_1^i, \dots, p_{m_i}^i\}$, let S^i be the set of terms generated from $\{p_1^i, \dots, p_{m_i}^i\}$, and $s \in S^i$ be an arbitrary term, then a probability value for s is defined as follows.

Definition 7 (A counterpart of P_E^{supp}) Let E be a knowledge profile and $s \in S^i$ be a term generated as above. Then the probability value of s is calculated by

$$P_E^{suppdep}(s) = \frac{Supp_E(s)}{\sum_{s \in S^i} Supp_E(s)}. \quad (5)$$

where $Supp_E(s)$ is defined similarly to $Supp_E(p)$ in Def 4 when replacing p with s .

The soundness of $P_E^{suppdep}$ is verified by $\forall S^i, \sum_{s \in S^i} P_E^{suppdep}(s) = 1$. Since a possible world w can be written as $w = \bigwedge_{i=1}^k s_i$ where $s_i \in S^i$ is a term and k is the number of partitions (partitioned by dependence relations) described above, and s_i is independent of s_j for $i \neq j$, we have $P_E^{suppdep}(w) = \prod_{i=1}^k P_E^{suppdep}(s_i)$.

Example 7 Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{(p \vee q) \wedge r\}$, $K_2 = \{\neg p \wedge \neg r\}$ and $K_3 = \{q\}$. We further assume that p, q are dependent, and r is independent of p and q ,

¹Determining whether atoms are (in)dependent can be a complex task which is beyond the scope of this paper. Here we assume that (in)dependence relations among atoms concerned are given.

which gives a partition of atoms as $\mathcal{L} = \{\{p, q\}, \{r\}\}$. Then we have $S^1 = \{p \wedge q, p \wedge \neg q, \neg p \wedge q, \neg p \wedge \neg q\}$, $S^2 = \{r, \neg r\}$ and

$$\begin{aligned} P_E^{suppdep}(p \wedge q) &= 1/3, & P_E^{suppdep}(p \wedge \neg q) &= 1/9, \\ P_E^{suppdep}(\neg p \wedge q) &= 4/9, & P_E^{suppdep}(\neg p \wedge \neg q) &= 1/9, \\ P_E^{suppdep}(r) &= 5/9, & P_E^{suppdep}(\neg r) &= 4/9. \end{aligned}$$

Now we can calculate probability values on possible worlds and also on formulae. For instance, we have $P_E^{suppdep}(p \wedge q \wedge r) = 5/27$ and $P_E^{suppdep}(p \wedge r) = 20/81$.

Proposition 7 Let E be a knowledge profile, then for any $p \in \mathcal{F}_E$, we have $P_E^{supp}(p) = P_E^{suppdep}(p)$.

This proposition shows that $P_E^{suppdep}$ is actually a counterpart of P_E^{supp} and hence of P_E^{IndK} .

Propositions 4 and 7 together reveal that no matter how atoms are related (dependent or independent), the support to them by the original knowledge bases in a profile is solely determined by these KBs. Therefore, all these apparently different probability functions produce the same probability values on atoms, but with different probability values on formulae giving different (in)dependence assumptions. This is not surprising, because (in)dependence relations only influence the calculation of probability values on formulae (and possible worlds) but not on atoms.

Now we consider the counterpart of P_E^r .

Definition 8 (A counterpart of P_E^r) Let E be a knowledge profile and $s \in S^j$ be a term generated as above. Then the probability value of s is calculated by

$$P_E^{rdep}(s) = \frac{1}{|E|} \sum_{i=1}^{|E|} \frac{Supp_{K_i}(s)}{\sum_{s \in S^i} Supp_{K_i}(s)}. \quad (6)$$

where $Supp_{K_i}(s)$ is defined similarly to $Supp_{K_i}(p)$ in Def 4 when replacing p with s .

Similarly, for P_E^{rdep} , we have $\forall S^i, \sum_{s \in S^i} P_E^{rdep}(s) = 1$.

Example 8 Let $E = \{K_1, K_2, K_3\}$ such that $K_1 = \{(p \vee q) \wedge r\}$, $K_2 = \{\neg p \wedge \neg r\}$ and $K_3 = \{q\}$, and we have p, q are dependent, and r is independent of p and q , then we have

$$\begin{aligned} P_E^{rdep}(p \wedge q) &= 5/18, & P_E^{rdep}(p \wedge \neg q) &= 1/9, \\ P_E^{rdep}(\neg p \wedge q) &= 4/9, & P_E^{rdep}(\neg p \wedge \neg q) &= 1/6, \\ P_E^{rdep}(r) &= 1/2, & P_E^{rdep}(\neg r) &= 1/2. \end{aligned}$$

We also have the following result.

Proposition 8 Let E be a knowledge profile, then for any $p \in \mathcal{F}_E$, we have $P_E^r(p) = P_E^{rdep}(p)$.

A possible world $w = p'_1 \wedge \dots \wedge p'_n$ (where p'_i is p_i or $\neg p_i$) can be rewritten as $w = s_1 \wedge \dots \wedge s_k$ (where each s_i is a term from subset L_i of a partition). Since atoms in different subsets of a partition are mutually independent, we have $P_E^{rdep}(w) = \prod_{i=1}^k P_E^{rdep}(s_i)$ (or $P_E^{suppdep}(w) = \prod_{i=1}^k P_E^{suppdep}(s_i)$).

Definitions 7 and 8 together can be regarded as a general framework subsuming Definitions 1 and 2 respectively when no independence relationships are observed, and Definitions 5 and 6 respectively when all the atoms are independent. Formally, we have the following two propositions.

Proposition 9 Given a knowledge profile E , if no independence relationships on atoms are given, then $P_E^{suppdep}$ (Eq. 5) is reduced to P_E^{IndK} (Eq. 1); if all atoms are pair-wise independent, then $P_E^{suppdep}$ is reduced to P_E^{supp} (Eq. 3).

Proposition 10 Given a knowledge profile E , if no independence relations are specified, then P_E^{rdep} (Eq. 6) is reduced to P_E^{IndE} (Eq. 2); if all atoms are assumed pair-wise independent, then P_E^{rdep} is reduced to P_E^r (Eq. 4).

From Propositions 9 and 10, and the fact that P_E^{IndK} and P_E^{IndE} follow the principles of indifference on knowledge bases and on profile, respectively, we can say that $P_E^{suppdep}$ (resp. P_E^{supp}) and P_E^{rdep} (resp. P_E^r) also follow the principles of indifference on knowledge bases and on profile, respectively, except that the principles are focused on terms (resp. atoms) instead of on possible worlds. Therefore, the principle of indifference is the foundation for defining all the probability distributions.

Related Work

In (Konieczny and Pino-Pérez 1998), an merging operator Δ is a mapping from knowledge profiles to knowledge bases satisfying the following set of postulates:

- A1 $\Delta(E)$ is consistent.
- A2 If E is consistent, then $\Delta(E) = \bigwedge E$.
- A3 If $E_1 \leftrightarrow E_2$, then $\vdash \Delta(E_1) \leftrightarrow \Delta(E_2)$.
- A4 If $K \wedge K'$ is not consistent, then $\Delta(K \sqcup K') \not\vdash K$.
- A5 $\Delta(E_1) \wedge \Delta(E_2) \vdash \Delta(E_1 \sqcup E_2)$.
- A6 If $\Delta(E_1) \wedge \Delta(E_2)$ is consistent, then $\Delta(E_1 \sqcup E_2) \vdash \Delta(E_1) \wedge \Delta(E_2)$.

In our framework, an operator is a mapping from knowledge profiles to probability distributions. Therefore, it is nature to consider which postulates given above may be applicable to the operators in our merging framework. Since independence relationships are not considered in (Konieczny and Pino-Pérez 1998), here we only consider operators producing P_E^{IndK} and P_E^{IndE} as merging results. To facilitate the comparisons, we first define the set of most plausible possible worlds generated based on either the total pre-order P_E^{IndK} on W or P_E^{IndE} on W . To this, we use $max(W, \leq)$ to denote the set $\{w \in W \mid \nexists w' \in W, w' \geq w\}$ where \leq is a total-preorder, and $form(\{w_1, \dots, w_k\})$ to denote a propositional formula whose models are w_1, \dots, w_k .

Definition 9 Let E be a knowledge profile, then the sets of most plausible possible worlds from operator Δ^{IndK} based on P_E^{IndK} and operator Δ^{IndE} from P_E^{IndE} are defined respectively as
 $\Delta^{IndK}(E) = form(max(W, \leq_{P_E^{IndK}}))$ and
 $\Delta^{IndE}(E) = form(max(W, \leq_{P_E^{IndE}}))$.

Then, we have the following result.

Proposition 11 Operator Δ^{IndK} satisfies A1-A6, and operator Δ^{IndE} satisfies A1-A3, A5, A6.

Due to space limitation, we omit the comparison between $\Delta^{IndK}(E)$ and $\Delta^{IndE}(E)$ with the Majority and the Arbitrary postulates in (Konieczny and Pino-Pérez 1998). We can conclude that both $\Delta^{IndK}(E)$ and $\Delta^{IndE}(E)$ are neither majority operators nor arbitrary operators.

Conclusion

In this paper, we have investigated how probability distributions can be obtained when merging multiple knowledge bases. A key motivation for obtaining a probability distribution rather than a flat knowledge base is to better preserve inconsistent knowledge from the original KBs. We have proposed two definitions to define an induced probability distribution based on two extensions to the principle of indifference to possible worlds proposed in (Bacchus et al. 1996).

These two definitions were then revised for situations where all atoms are independent or some atoms are dependent, since independence relations between atoms can significantly affect induced probability distributions. Comparison with logic based merging postulates shows that our merging framework is consistent with logic based merging when no information about (in)dependent relationships among atoms is available.

One possibility for future work is to investigate if other approaches to inducing probability distributions can be developed when the requirement on principle of indifference on possible worlds (resp. KBs, or profile) is removed.

Acknowledge: This research is supported by the UK EPSRC projects EP/D070864/1 and EP/D074282/1.

References

- Bacchus, F.; Grove, A.; Halpern, J.; and Koller, D. 1996. From statistical knowledge bases to degrees of belief. *Artificial Intelligence* 87(1-2):75–143.
- Baral, C.; Kraus, S.; Minker, J.; and Subrahmanian, V. 1992. Combining knowledge bases consisting of first-order theories. *Computational Intelligence* 8(1):45–71.
- Benferhat, S.; Lagrue, S.; and Rossit, J. 2007. An egalitarian fusion of incommensurable ranked belief bases under constraints. In *Proc. of AAAI'07*, 367–372.
- Delgrande, J.; Dubois, D.; and Lang, J. 2006. Iterated revision as prioritized merging. In *Proc. of KR'06*, 210–220.
- Knight, K. 2002. Measuring inconsistency. *Journal of Philosophical Logic* 31(1):77–98.
- Konieczny, S., and Pino-Pérez, R. 1998. On the logic of merging. In *Procs. of KR'98*, 488–498.
- Konieczny, S.; Lang, J.; and Marquis, P. 2004. Da^2 merging operators. *Artif. Intel.* 157:49–79.
- Konieczny, S. 2000. On the difference between merging knowledge bases and combining them. In *Proc. of KR'00*, 12–17.
- Paris, J. 1994. *The Uncertain Reasoner's Companion*. Cambridge Univ. Press.
- Qi, G.; Liu, W.; and Bell, D. A. 2006. Merging stratified knowledge bases under constraints. In *Proc. of AAAI'06*, 281–286.
- Revesz, P. 1997. On the semantics of arbitration. *Internat. J. Algebra Comput.* 7(2):133–160.