Development of a novel asset management system for power transformers based on Ontology

Mr Long Yan
Department of Electrical Engineering and Electronics
University of Liverpool
Liverpool
Email: l.yan1@liv.ac.uk

Dr Wenhu Tang
Department of Electrical Engineering and Electronics
University of Liverpool
Liverpool
Email: whtang@liv.ac.uk
Telephone: (+44)1517944615

***Abstract*—This paper aims to develop a formalised framework, which can perform reasoning with uncertainty in semantic web, for adopting rules with interlinked relationships to form an interoperable knowledge base for power transformers and developing a probabilistic diagnosis system to provide quantified confidence support if uncertainties occur. The framework provides a set of structural translation rules to map an OWL taxonomy into a Bayesion Network (BN) directed acyclic graph. Firstly, the essential concepts of BNs are introduced, which are graphical representations of uncertain knowledge. The algorithms of knowledge integration is used to refine an existing BN with more reliable sources. Secondly, the advantages and shortages of crisp logic based ontology are introduced. The framework augments and supplements OWL with additional functions for representing and reasoning with uncertainty based on BN. Finally, an example of transformer diagnosis is provided, which shows the existing BN has been refined by new constraints.**

Index Terms*—Bayesion Network, Web Ontology Language, Iterative Proportional Fitting Procedure, Transfomer

I. Introduction and Motivation

As a crucial device, accurate transformer incipient fault diagnosis can extend the useful life of power transformers and further increase grid reliability to avoid power blackouts. In industrial practice, dissolved gas analysis (DGA) is a very efficient tool for such purpose, since it can warn about an impendent problem, provide an early diagnosis, and ensure a transformer's maximum uptime. The core tasks of transformer fault diagnosis are to identify the relationships between fault gas ratios and fault types [1]. It is actually a probabilistic reasoning process to compute the unknown probabilities of a certain type of fault, given new evidence about groups of key gas ratios based upon known probabilities. In recent research, Bayesian Network (BN) has been chosen as one of the most effective and accurate methods to deal with such issues.

A. Bayesion Network

Over the last decade, BN has become an increasingly important method for uncertainty reasoning. Briefly, it contains a set of nodes and links, which allows us to represent and reason about an uncertain domain. Most commonly, BNs are considered to be representation of joint probability distributions. A particular node is conditional only on the values of its parent nodes. These conditional dependencies between a particular node and its parent nodes are represented by Conditional Probability Tables (CPTs). In practice, most publications only focus on how to apply BN on power transformer diagnosis. In real diagnosis, the number and type of signal collection increase dramatically with the increasing of the system scale and complexity, like temperature, vibration, noise, discharging and so forth. Knowledge (*e.g.* DGA ratio or related parameters) can be collected from different sources. They are sometimes mutually related. However, the knowledge from different sources can also be conflict. If more reliable probabilistic knowledge is obtained, an existing BN can be merged with the reliable probabilistic knowledge rather than be reconstructed over the new situation. The procedure to build a BN and construction of CPTs are complex and time consuming. Thus, to find a method to modify CPTs of a BN to meet a set of given low dimensional probabilistic constraints becomes very important. In other words, we aim to seek an algorithm, which can be used to design new BNs, to merge small BNs into a large one, and to refine an existing BN with new or more reliable probabilistic information.

B. Knowledge integration

Normally, the general probabilistic knowledge integration problem can be solved by means of iterative procedures. The best-known example is the Iterative Proportional Fitting Procedure (**IPFP**), which was firstly proposed by Kruithof [2], iterating constraints based on the minimum K-L divergence, so as to realise the probabilistic constraint satisfaction. It was proven by Csizsár [3] that **IPFP** converges if the input set of low-dimensional probability distributions is consistent, i.e., if there exists a probability distribution whose marginals equal to the probability distributions from the input set. **IPFP** is only appropriate for the knowledge base, which is joint probability distribution, and it cannot be applied directly onto BN. To solve this issue, an extension of **IPFP**known as **E-IPFP** has been proposed [4], which iterates and changes the net arguments to encode the given probabilistic knowledge to CPTs of BN. Each step of iteration of both **IPFP** and **E-IPFP** is proceeded in the whole joint probability distribution, and the joint probability distribution of the resulting network will be as close as possible to that of the original network.

C. Web Ontology Language (OWL)

The idea of semantic web was started in 1998, aiming to add a layer of machine-understandable information over the existing web data to provide meaning or semantics to these data [6]. The core of the semantic web is "ontology". The most common concept of ontology is defined as an explicit specification of a conceptualization by Thomas [7]. It is a formal representation of the entities and relationships that can exist in a domain of application. It provides the common understanding of a domain, as well as the explicit relationship between the terms in different hierarchy. It has been proven that ontology knowledge is of great importance for interconnection and construction of large-scale software system to deal with power system related topics in many previous researches [5]. It is a good way to provide conceptual specific description and build the foundation for knowledge sharing. This paper focused on one of the ontology language, named Web ontology language (OWL) [6]. It is an emerging standard recommended by W3C, which is based upon description logics, providing decidable and sound inference mechanism. Classes, properties, axioms, and individual descriptions are defined in OWL. However, ontology languages are based on crisp logic, which cannot handle incomplete or partial knowledge about an application domain. Even if "concept A is a subclass of B" is True, the closeness of A and B can not be ensured. None of the existing ontology language including the most advanced OWL, provides a means to capture uncertainty about the concepts, properties and instances in a domain.

In our previous research, it has been proven that IEEE/IEC DGA coding scheme can be directly mapped into a BN solution, and this approach overcomes the drawbacks of missing codes in [1]. However, the previous research only focused on the use of BN approach on DGA. The procedure and calculation to build a BN is complex and time consuming. As discussed before, if reliable knowledge is obtained, it is inconvenient to reconstruct the BNs over the new situation. Here, to find an algorithm to modify CPTs of a BN to meet a set of given low dimensional probabilistic constraints becomes very important. Considering the benefit and drawbacks of ontology language, we aim to integrate BN in uncertainty reasoning extended OWL with probability description, so as to let it support uncertainty knowledge and incomplete information, thus we can preserve the advantages of both. A domain ontology, which was extended by probability information in Protégé, also was built [4]. In general, this paper aims to construct a framework which augments and supplements OWL with additional expressive power for representing and reasoning with uncertainty based on modified BNs. Furthermore, this framework can be utilised to construst a small power transformer diagnosis system.

II. METHODS

A. Structural translation

A transformer winding is always covered with paper, and located in oil. The insulation problem may happen, when trace water occurs whether in the oil or paper. Both phenomena can be regarded as a relationship of union. OWL can be used to define these concepts and relations. This structure is mapped onto a binary variable node in the translated BN using Jena Java API. There are two kinds of OWL file. One is used for defining the relationship among concepts as shown in List 1, the other is used for defining the values of concepts (probability) in List 2. For the probability OWL, a class named c has been defined, who contains two states, namely True ($P(c) = 0.3$) and False. Applying Jena API, every defined concept in the OWL file is mapped onto a binary variable node in the translated BN as shown in Figure **??**. Besides the concept nodes (C-nodes) $e.g. Insulation$, the translated BN contains another kind of node named Logical nodes (L-nodes) $e.g. LNodeUnion$. Briefly, C-nodes are pure concept subsumption hierarchy, which can be easily translated into BN based on subclass relations in OWL; while L-nodes are their logical relation. The L-nodes bridge concept nodes that are associated by logical relations, and they are leaf nodes, with only in-arcs. L-nodes help avoid forming cycles in translated BN. In this project, we only consider 5 types of L-nodes, namely union, equivalent, disjoint, intersection and complement.

To complete the translation, the remaining issue is to assign a conditional probability table to each variable node. To approach the CPT for C-nodes and L-nodes are different. CPT for an L-node can be determined by the logical relation it represents so that when its state is "True", the corresponding logical relation holds among its parents. Table I is the CPT of logic operator "LNodeUnion" in the insulation problem.

Listing 1: OWL file for relations

```
<owl:Class rdf:ID="water">
    <rdfs:subClassOf rdf:resource="#insulation"/>
        <owl:unionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#water_in_oil"/>
        <owl:Class rdf:about="#water_in_paper"/>
    </owl:unionOf>
</owl:Class>
```

Listing 2: OWL file for Probabilities

```
<Variable rdf:ID="c">
     <hasClass>C</hasClass>
     <hasState>True</hasState>
</Variable>
<PriorProb rdf:ID="P(c)">
     <hasVariable>c</hasVariable>
     <hasProbValue>0.3</hasProbValue>
</PriorProb>
```

TABLE I: CPT for LNodeUnion

C1	C2	C	True	False
T	T	T	1	0
T	T	F	0	1
T	F	T	1	0
T	F	F	0	1
F	T	T	1	0
F	T	F	0	1
F	F	T	0	1
F	F	F	1	0

The logical relations defined in the original ontology will be held in the translated BN, making the BN consistent with the OWL semantics, if state of L-nodes are all set to be "True". For the C-nodes, they are conditionally dependent with this situation, the JPD of all C-nodes in this subspace is consistent with all the given prior and conditional probabilities attached to the nodes in C-nodes. It becomes a problem of modifying the existing BN with given probabilistic constraints.

B. Preliminaries

According to probability theory frame, the domain problem can be represented by a set of random variables:

$$X = (X_1, X_2, ..., X_n)$$

where $P(X) = P(X_1, X_2, ..., X_n)$ denotes an n-dimensional joint probability distribution if for every assignment or instantiation $x = (x_1, x_2, ..., x_n) \in X$,$0 \leq P(X) = P(X_1, X_2, ..., X_n) \leq 1$ and $\sum\limits_{x \in X} P(X = x) = 1$ as x runs through all possible assignments of X

*Definition 1 (**Probabilistic Constraints**):* The new probabilistic knowledge set is defined as a low dimensional probabilistic constraints, which is a subset of X, denoted as:

$$R = \{R(Y^1), R(Y^2), ..., R(Y^m)\}$$

where $Y^j \subseteq X$, m is the number of constraints.

*Definition 2 (**Consistent Probabilistic Constraints**):* If exist one joint probability distribution $P(X)$, satisfy the probabilistic constraint set R, i.e. for each $0 \leq j \leq m$, $P(Y^j) = R(Y^j)$. Therefore, the R is consistent, otherwise, R is inconsistent. Constraints (probabilistic knowledge) come from different sources, which leads to the set of constraints are not consistent, even they are contradictory. For example, as we know the relationship between woman and man should be complementation, i.e. $P(Man) + P(Woman) = 1$. However, the probability for each concept we obtained from different sources can be as these form: $P(Man|Animal, Human) = 0.564$ and $P(Woman|Animal, Human) = 0.664$. It is impossible to find a JPD to satisfy both probabilistic constraints. Therefore, the constraint set is inconsistent, and **IPFP** cannot be used in this situation. At current stage, we only consider the consistent situation.

*Definition 3 (**Kullback-Leibler Divergence**):* It is also known as K-L distance or relative entropy, which is a measure to reflect the difference between two joint probability.

Let P be the set of joint probability distributions (JPDs) over random variables $X = (X_1, X_2, ..., X_n)$, and $Q, Q^* \in P$. The K-L divergence is defined as following:

$$I(Q^* \| Q) = \sum_{x \in X, Q^*(x) > 0} Q^*(x) \log_2 \frac{Q^*(x)}{Q(x)}$$

We set $0 \cdot \log \frac{0}{Q} = 0, Q^* \cdot \log \frac{Q^*}{0} = \infty$

*Definition 4 (**Probabilistic Knowledge Integration**):* Basically, it is the process to seek new joint probability distribution, which satisfies given probabilistic constraints.

i.e. Given joint probability distribution $Q(X)$ and probabilistic constraints $R = \{R(Y^1), R(Y^2), ..., R(Y^m)\}$, construct new JPD $Q^*(X)$, which satisfies R. Furthermore, the new JPD $Q^*(X)$ has minimum K-L divergence with initial JPD $Q(X)$.

*Definition 5 (**I-projection**):* Probability distributions from a given set minimizing K-L divergence with respect to a given distribution are called I-projection.

$$I(Q^* \| Q) = \min I(Q^* \| Q)$$

It can be defined as the formula below:

$$Q^*(X) = \begin{cases} 0 & if \quad Q(Y) = 0 \\ Q(X) \cdot \frac{R(Y)}{Q(Y)} & if \quad Q(Y) > 0 \end{cases}$$

Normally, the constraint set contains more than one constraint, therefore, the problem become to a iterative process. In other words, repeatedly use the constraints in order for each iterative step until the result converges.

*Definition 6 (**IPFP**):* is a procedure for determining a joint distribution, which satisfies all constraints by repeating the following computational process:

1) Initial state:
   $Q_0(X)$, $R = \{R(Y^1), R(Y^2), ..., R(Y^m)\}$
2) For k = 1, repeatedly do the following iterative process until the result converges:

   a) $i = ((k-1) \bmod m) + 1$

   b)

$$Q_{(k)}(X) = \begin{cases} 0 \\ \quad if \quad Q_{(k-1)}(Y^i) = 0 \\ Q_{(k-1)}(X) \cdot \frac{R_i(Y^i)}{Q_{(k-1)}(Y^i)} \\ \quad if \quad Q_{(k-1)}(Y^i) > 0 \end{cases}$$

3) k = k + 1

*Definition 7 (**C-IPFP**):* If the probabilistic constraint is conditional probability distribution rather than marginal probability, **IPFP** can be developed to **C-IPFP** [8] [9], which is suitable to deal with the constraint in the form of $R(Y|Z)$, where Y is conditional to Z.
The algorithm can be concluded as:

1) Initial state: $Q_0(X)$, $R = \{R_1, R_2, ..., R_m\}$
2) For k = 1, repeatedly do the following iterative process until the result converges:

   a) $i = ((k-1) \bmod m) + 1$

   b)

$$Q_{(k)}(X) = \begin{cases} 0 \\ \quad if \quad Q_{(k-1)}(Y^i | Z^i) = 0 \\ Q_{(k-1)}(X) \cdot \frac{R_i(Y^i|Z^i)}{Q_{(k-1)}(Y^i|Z^i)} \\ \quad if \quad Q_{(k-1)}(Y^i | Z^i) > 0 \end{cases}$$

3) k = k + 1

Example Bayesian Network N_0 consists four binary variables $\{A, B, C, D\}$, and its JPD can be calculated by chain rule as shown in Table IIa. Here, "1" is used for a "True" state and "0" is used for a "False" state. If the probabilistic constraint set is $R(A, D) = (0.1868, 0.2132; 0.1314, 0.4686)$, apply **IPFP** to the JPD Q_0 of BN N_0. The fourth iterative results has been chosen as an example, and the resultant JPD Q_4 is shown in Table IIb.

Variables				
A	B	C	D	Prob.
1	1	1	1	0.0048
1	1	1	0	0.0432
1	1	0	1	0.0272
1	1	0	0	0.0048
1	0	1	1	0.0864
1	0	1	0	0.1056
1	0	0	1	0.0896
1	0	0	0	0.0384
0	1	1	1	0.0126
0	1	1	0	0.1134
0	1	0	1	0.1989
0	1	0	0	0.0351
0	0	1	1	0.0378
0	0	1	0	0.0462
0	0	0	1	0.1092
0	0	0	0	0.0468

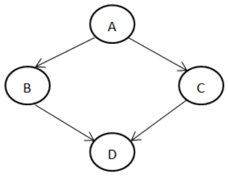
(a) Original JPD

Variables				
A	B	C	D	Prob.
1	1	1	1	0.0043
1	1	1	0	0.0480
1	1	0	1	0.0244
1	1	0	0	0.0053
1	0	1	1	0.0776
1	0	1	0	0.1173
1	0	0	1	0.0805
1	0	0	0	0.0426
0	1	1	1	0.0046
0	1	1	0	0.2200
0	1	0	1	0.0729
0	1	0	0	0.0139
0	0	1	1	0.0897
0	0	1	0	0.0400
0	0	0	1	0.0206
0	0	0	0	0.0908

(b) Modified JPD with one constraints

From the JPD Q_4, the CPTs of the four variables are extracted from Q_4. (Figure 1) It is easy to prove that JPD Q_4 satisfies $R(A, D)$, but $Q_4(A, B, C, D) \neq Q_4(A) \cdot Q_4(B|A) \cdot Q_4(C|A) \cdot Q_4(D|B, C)$. This example shows that **IPFP** can not be applied to BN directly. BN has two important factors. Firstly, it is a directed acyclic graph, which is denoted by G. Secondly, It must obey the chain rule: $Q(X_1, ..., X_n) = \prod_{i=1}^{n} P(X_i | \pi_i)$, where π_i denotes all the parent nodes of X_i. In this case, we treat the requirement as a probabilistic constraint $R = \prod_{i=1}^{n} Q_{(k-1)}(X_i | \pi_i)$ in **IPFP**. Here $Q_{(k-1)}(X_i | \pi_i)$ are extracted from $Q_{k-1}(X)$.We call the constraint structural constraint.

A	
1	0
0.4	0.6



A	B 1	B 0
1	0.2051	0.7949
0	0.6094	0.3906

A	C 1	C 0
1	0.6178	0.3822
0	0.5469	0.4531

B	C	D 1	D 0
1	1	0.0322	0.9678
1	0	0.57	0.43
0	1	0.3065	0.6935
0	0	0.4745	0.5255

Fig. 1: Running IPFP with $R(A, D)$

Consider a BN (N) on a set of n variables $X = (X_1, X_2, ..., X_n)$ with distribution $Q(X)$ and a set of consistent constraints R, find N^* that meets the following three requirements:

1) $G_0 = G^*$ (Both networks have the same structure);
2) $Q^*(X)$, the distribution of N^*, satisfies all constraints in R;
3) K-L divergence $I(Q^* \| Q)$ is miminmum among all distributions that meet requirements 1 and 2.

*Definition 8 (**E-IPFP**):* is a simple extension of **IPFP** by including the structural constraint as the $(m+1)^{th}$ constraint $R_{m+1}(X)$. The algorithm of **E-IPFP** is shown as following:

1) Initial state: $Q_0(X)$, $R = \{R_1, R_2, ..., R_m\}$
2) Starting with k = 1, repeat following iterative process until the result converges:
   a) $i = ((k-1) \bmod (m+1)) + 1$
   b) if $i < m+1$
      i) if $(R_i \in R_m)$ (Using IPFP for marginal constraints)
      $$Q_{(k)}(X) = Q_{(k-1)}(X) \cdot \frac{R_i(Y^i)}{Q_{(k-1)}(Y^i)}$$
      ii) else if $(R_i \in R_c)$ (Using C-IPFP for conditional constraints)
      $$Q_{(k)}(X) = Q_{(k-1)}(X) \cdot \frac{R_i(Y^i | Z^i)}{Q_{(k-1)}(Y^i | Z^i)}$$
   c) else extract $Q_{(k-1)}(X_i | \pi_i)$ from $Q_{(k-1)}(X)$ according to G_0;
   $$Q_{(k)}(X) = \prod_{i=1}^{n} Q_{(k-1)}(X_i | \pi_i)$$
   d) k = k + 1
3) return $N^*(X)$ with $G^* = G_0$

III. An Example

To utilise **E-IPFP** on power transformer diagnosis, the example in section II-A will be introduced in detail in this part. An OWL file named PT.owl has been created by protégé. This ontology defines the following six power transformer related concept classes and several logical relations among these concepts: "insulation" is regarded as a primitive concept class; "Water", "Insulation in oil" , and "Insulation in paper" are subclasses of "Insulation"; "water in oil" and "water in paper" are subclasses of "water"; "water in oil" is an intersection of "water"and "Insulation in oil"; "water in paper" is an intersection of "water" and "Insulation in paper"; "water" is the union of "water in paper" and "water in oil". Therefore, there are four L-Nodes within the six concept classes, and a set of probabilistic constraints are also provided in OWL file PTprob.owl as following:

1) $P(Insulation) = 0.5$;
2) $P(Insulation\ in\ oil | Insulation) = 0.48$;
3) $P(Insulation\ in\ paper | Insulation) = 0.5$;
4) $P(water | Insulation) = 0.9$;
5) $P(water\ in\ oil | water) = 0.51$;
6) $P(water\ in\ paper | water) = 0.49$;

Figure 2 shows the mapped BN. When all L-nodes are set to be True, the concept nodes are set to the corresponding values as provided above. The initial CPTs and final CPTs, which are generated by E-IPFP have been listed in TableII. The initial states can be set to any arbitrary values between 0 and 1. In this example, values on the first row are set to 0.5. After the process of **E-IPFP**, all the CPTs have been changed.

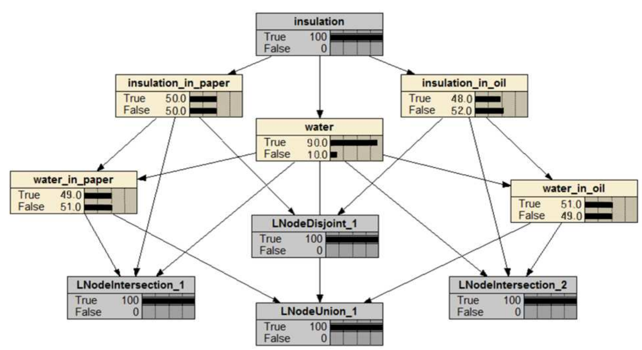


Fig. 2: BN obtained from OWL

IV. Discussion and Future Work

In this example, the initial states, which can be regarded as a primitive BN, have no influence on the resultant CPTs. The constraint sets are regarded as the new

TABLE II: Initial and Final CPTs of BN

Insulation			
Initial		Final	
T	F	T	F
0.5	0.5	0.927	0.073

Insulation	Insulation in Paper			
	Initial		Final	
	T	F	T	F
T	0.5	0.5	0.956	0.044
F	0	1	0	1

Insulation	Insulation in Oil			
	Initial		Final	
	T	F	T	F
T	0.5	0.5	0.955	0.045
F	0	1	0	1

Insulation	Water			
	Initial		Final	
	T	F	T	F
T	0.5	0.5	0.823	0.187
F	0	1	0	1

Insulation in Oil	Water	Water in Oil			
		Initial		Final	
		T	F	T	F
T	T	0.5	0.5	0.514	0.486
T	F	0	1	0	1
F	T	0	1	0	1
F	F	0	1	0	1

Insulation in Paper	Water	Water in Paper			
		Initial		Final	
		T	F	T	F
T	T	0.5	0.5	0.471	0.529
T	F	0	1	0	1
F	T	0	1	0	1
F	F	0	1	0	1

conditions from more reliable sources. The computation of **E-IPFP** can be regarded as a process to refine a BN. Power transformer always contains a lot of components and symptoms. The network obtained will be extremely large and complex. As mentioned before, the procedure to build a BN is complex and time consuming. **E-IPFP** is a good way to refine and adjust the primitive BN with other constraints. It has been proven that **E-IPFP** performs better than simulated annealing (SA) and genetic algorithm (GA) for constructing CPTs of regular nodes in the translated BN [10]. As the knowledge of Ontology is getting more and more widely used in power system, some necessary extension can be attached to it. This framework augments and supplements OWL with additional expressive power for representing and reasoning with uncertainty based on BN. This paper presented a small power system related example, which utilised the extended OWL to improve the accuracy of reasoning. There are still some advanced topics worthy of study in the future. **E-IPFP** manipulates the CPTs through the entire joint probability distribution. If the network is very large, the computation process increases dramatically. It is expected to decompose **E-IPFP** into smaller scale. As mentioned in Section II-B, constraints are from different sources. It is of great importance to consider inconsistent or incomplete input sets in the future research.

REFERENCES

[1] W.H.Tang, Z.Lu, and Q.H.Wu, “A Bayesian network approach to power system asset management for transformer dissolved gas analysis,” University of Liverpool, 2005.

[2] R. Kruithof, “Telefoonverkeersrekening,” De Ingenieur, vol. 52, pp. E15-E25, 1937

[3] D.T.Brown, “A note on approximations to discrete probability distributions”. *Infrom. and control*. 2:386 392.1959

[4] Z. Ding, Y. Peng, R. Pan, and Y. Yu, “A bayesian methodology towards au- tomatic ontology mapping,” in Proceedings of the 1st International Workshop on Contexts and Ontologies: Theory, Practice and Applications at AAAI-05, Pittsburgh, PA, July 2005.

[5] Q.Ma, J.Z.Guo, and Y.H.Guo, “An ontology for power system operation analysis,” *Electric Utility Deregulation, Restructuring and Power Technologies, vol. 2, pp.597-601, 2004*

[6] Semantic Web Homepage: http://www.w3.org/2001/sw/

[7] T.R. Gruber, “Toward principles for the design of ontologies used for knowledge sharing,” *Int. J. Human-Computer Studies*, November 1995, 43, 907–928.

[8] H.H.Bock, “A conditional iterative proportional fitting (CIPF) algorithm with applications in the statistical analysis of discrete spatial data,” *Bull. ISI, Contributed Papers of 47th Session in Paris*, vol.1,pp.141-142,1989.

[9] E.Cramer, “Probability measures with given marginals and conditionals: I-projections and conditional iterative proportional fitting,” *Statistics and Decisions, vol. 18, pp. 311-329, 2000*

[10] Z. Ding, Y. Peng, and R. Pan, “A Bayesian approach to uncertainty modelling in OWL ontology,”*Information for the defense community*.