Appl. Math. Inf. Sci. **9**, No. 3, 1611-1621 (2015)

Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.12785/amis/090358

Intelligent Expert Systems for Location Planning

Daizhong Tang[1], *Jiangang Shi*[1] *and Wei Wang*[2,3,*]

[1] School of Economics and Management, Tongji University, Shanghai, 200092, China
[2] Department of Computer Science and Engineering, Tongji University, Shanghai 200092, China
[3] Key Laboratory of Embedded System and Service Computing, Ministry of Education, Shanghai 200092, China

Received: 28 Aug. 2014, Revised: 29 Nov. 2014, Accepted: 30 Nov. 2014
Published online: 1 May 2015

Abstract: We discuss and illustrate how Semantic Web technologies can support information integration and create semantic mashups. An intelligent expert system for location planning is presented to show the efficiency of our method. Through the ontology of location planning, the expert system allows the integration of heterogeneous online location information. An integrated knowledge process is developed to guarantee the whole engineering procedure. Based on Bayesian network technique, the system recommends well planed attractions to a user by taking into account the behaviour both of the user and of other users.

Keywords: Expert System, Location Planning, Bayesian network

1 Introduction

There is a common perception that there are two competing visions about the future evolution of the Web: the Semantic Web and Web 2.0. The technologies and core strengths of these visions are complementary, rather than in competition. In fact, both technologies need each other in order to scale beyond their own strongholds.

Web 2.0 has enabled contributions to the Web development on an unprecedented scale, through simple interfaces that provide engaging interactions. This wealth of data has spawned countless mashups that integrate heterogeneous information, but using techniques that will not scale beyond a handful of sources [1]. In contrast, the Semantic Web aims to provide a new framework that can enable knowledge sharing and reusing. Semantic Web uses agent technology, ontology, and a number of standard markup languages, such as RDF, to formally model information represented in web resources. It can learn from Web 2.0's focus on community and interactivity, while Web 2.0 can draw from the Semantic Web's rich technical infrastructure for exchanging information across application boundaries.

This paper illustrates how Semantic Web technologies can support information integration and make it easy to create semantic mashups (semantically integrated resources). We proposed an expert system based on intelligent ontologies. Through these ontologies of location planning, the system allows integration of heterogeneous online location planning information. In addition, we have developed an integrated knowledge process to guarantee the whole knowledge engineering procedure. Based on Bayesian network technique method, the system recommends attractions to a user by taking into account the behaviors both of the user and of other users.

This paper makes the following contributions to the field of expert system and location planning.

– It proposes a method of building an intelligent ontology for location planning;
– It utilizes the powerful reasoning mechanism of Bayesian networks to solve key tasks in location planning;
– An integrated Bayesian knowledge engineering process is developed to assure the consistence and accuracy of the knowledge in support of semantic mashup.

Section 2 introduces some related works. Section 3 presents the details of our methods. In Section 4 we illustrate how to apply the proposed methods in support of building an intelligence tourism system. Section 5 concludes this paper.

* Corresponding author e-mail: willtongji@gmail.com

2 Related work

To our knowledge, no approaches that can be found in the literature cover all parts of the challenges that were derived in the previous section.

2.1 Location planning system

Recently personalized location recommendation systems have been gaining interest in tourism to assist travelers with location plans. A location plan consists of a number of stages, such as choosing destinations, selecting attractions, choosing accommodations, deciding routes, etc. However, at present most of the locations recommendations focus on the first stage - suggesting the destinations, with very few exceptions [7]. This paper is concerned with the second stage, which suggests a set of attractions in sequence at a given destination. Attractions, which are places intended to attract people to visit at a destination, are often the reason driving travelers to visit destinations [18,19,20].

2.2 Semantic Web

Past approaches to the topic of merging the Semantic Web with the existing world-wide web have often focused on the extraction of semantic content from HTML-based pages rather than the coexistence of HTML and RDF services as two different access mediums for the same data. MITs Piggy Bank [16] plug-in for Firefox uses pre-defined scrapers to extract semantic information out of the DOM structures of popular web sites such as Flickr and Amazon.com, for example [11]. Other examples use a combination of user-guided training and tree-based algorithms to learn how to scrape data from a site that presents multiple data objects of the same type with the same basic DOM layout [12].

Microformats [13] are a way for web scrapers and HTML developers to meet in the middle. Developers embed lightweight semantic markup into the class attribute of HTML elements. These markers provide a standard context through which to interpret the contents of the HTML tag. Semantic REST offers a different approach than the above tools because it focuses on the co-existence of HTML and RDF rather than a method for extracting RDF from HTML content. This different approach attempts to retain the capabilities and flexibility that make RDF and HTML attractive data formats in the first place.

2.3 Bayesian Networks with Knowledge Engineering

In recent years, recommendation assistants built around Bayesian Networks became especially popular [10]. The information necessary to create the models can be acquired from experts on system and design, as well as system technical documentation. Models can be developed entirely from repair records, or one can simply combine expert knowledge and data in the model development. The current literature on the use of BN models for recommendation is very broad. The BNs are applied in medicine, manufacturing, power generation, transportation, etc. Examples of the applications can be found in such works as [10].

On the other hand, although no existing methodology in doing knowledge engineering for service management, there is a great interest in using knowledge engineering method for building Bayesian Networks. Knowledge Engineering of Bayesian Networks or KEBN is an iterative or spiral approach to model prototype development, based on software development processes [16]. To date, the methodologies and associated support tools that accompany KEBN have been poorly developed.

3 Location Planning System Based on Semantic Mashup

3.1 Overview of the system

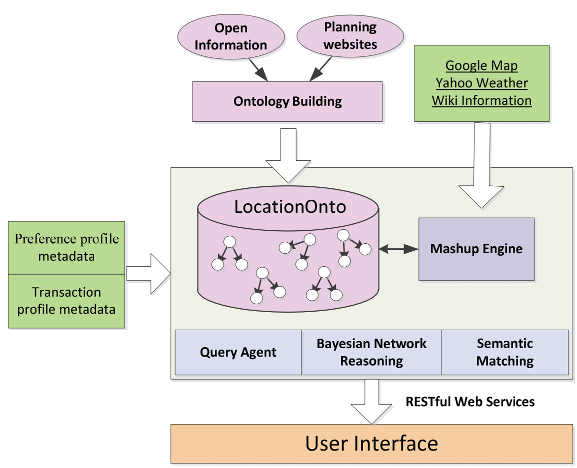

Fig. 1: Overview of the planning expert system architecture

The proposed expert system architecture is shown in Figure 1. Its divided into three parts. Firstly, it is the metadata consisting of preference profile and transaction profile. Secondly, the information repository of Ontology was build. Finally, the interface is the part for user query.

Preference profile is made by user profile such as name, age and address, etc., and transaction profile is made by information such as query, account of reference, which will be connected to in interface. In other words, it make the metadata to indicate users preference from user database and build user deposition database, also user preference profile is updated with the feedback of recommendation. For the user profile, we develop a Bayesian network to build user profiles as the representatives of usage patterns.

Information repository consists of travel ontology part and reasoning part. Planning Ontology is the construction domain ontology and we made rules in inference based on Bayesian networks and use them as the reasoning tools. We will present these two methods in the next section.

An ideal solution would bridge the two disciplines so that their access points aligned. Existing web sites could inter-operate with both OWL and RDF data while also serving as access points for semantically enabled clients or query architectures. We use the Representational State Transfer (REST) design methodology common amongst Web 2.0 developers today to support our implementation.

A popular theme in Web 2.0 is combining different Web services together to form a mashup. Distributing a query across a set of semantically enabled Web services as described above allows for the services to be combined into a semantic mashup.

The systems works in two steps: first the user fills in the interface so that his/her profile is generated, second the user states the question. Then, the system runs the context matching algorithm between the two ontologies and returns the answer as a text but also locating the proposed places/ points of interest on the map. The key characteristic of the ontology is that it is comprised of two steps. The first step, that of the design, concerns agreeing upon the main concepts of the ontology along with their properties. We include in the ontology not only these concepts that characterize/describe a tourist but also concepts that account for the personal information of the user respecting to his/her trip making. The building of the ontologies will be introduced in more detail in the next section.

3.2 Ontological model for location planning

3.2.1 Basic concept

Ontology, which is a theory about the nature of existence in philosophy, plays an important role in the integrating heterogeneous information. Various definitions of ontology have been given by researchers. The most commonly quoted definition of ontology is a formal, explicit specification of a shared conceptualization [14]. In other words, ontology describes a shared and common understanding of a domain that can be communicated among communities [7].

Concepts and relationships are basic components in ontology. Web documents are the most important source for deriving concepts and relationships. For this research, the online information of tourist attractions is the source to derive concepts and the relationships between these concepts in the tourist ontology.

3.2.2 Ontology development

Ontology is the central mechanism of the system. Tourist information and service resources are classified according to a common ontology to provide a common basis for searching, interpretation, and reasoning. As currently there is no existing commonly adopted ontology for tourism, the establishment of such ontology needs the expertise of experienced tourist consultants to incorporate the categorizations from different sub-domains, such as locations, tourist attractions, events, hotels, etc. Once the ontology framework is established, other automated method could be used to replace human effort in such kind of tedious process.

For example, since services like flight and hotel are relatively static and usually captured in existing systems, major ontology maintenance is mainly perform by importing from legacy system into the ontology. With further and wider adoption of XML and Web services, we can foresee that hotel and flight information could be gathered and updated automatically on demand in the near future.

Other services like local tour packages, attractions, dining, recreation, and entertainment are relative dynamic and more heterogeneous. As a result, a web-crawler is employed for handling this. In addition, we adopted WebXcript [18] to integrate these legacy tourist websites into the Web services. This is because we have no option to change the access methods of the existing agency websites over the Internet, and often cannot request for the provision of a programmatic interface. WebXcript simulates an interactive user accessing the target websites as well as communicate with the service providers with provision of Web services interface. The extracted information is stored in ontology or knowledge database for further reference. Further details of ontology development and maintenance are available in Chiu et al. [19].

3.2.3 Ontology engineering

By examining more than 80 websites for location planning in China and attractions at six cities including Beijing, Shanghai, Wuhan, Guangzhou, Xian and Chongqing, we found the main concepts for the ontology. This included terms used in the site map and the website menu. Web developers often group related contents into categories. The site structure is a common way for people to define a domain (e.g. travel).

We collected a number of websites that related to location planning from Yahoo! Directory. After removing duplicates, we were left with 232 websites. We visited each website and recorded the structural information.

There were 258 terms. After filtering and grouping similar terms. We further filtered and grouped together terms with similar meanings and used this information to model our upper-level travel ontology. In addition, we review WikiTravel website (Wikitravel is a Web-based collaborative travel guide project, based upon the wiki model) for more information. Up until now, there have 25,872 destination guides and other articles written and edited by Wikitravellers from around the globe.

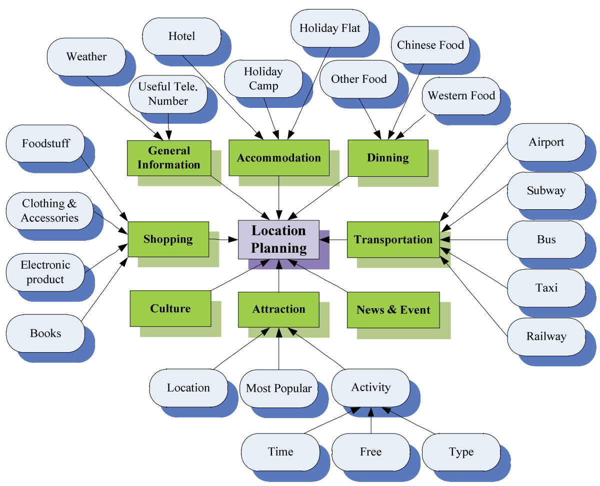

Fig. 2: Location Planning Ontology

To ensure the ontology would be reuse and share with others, we modeled travel ontology. Therefore, the ontology could be reused by different users in different regions. After collecting and analyzing the structural information, we defined travel ontology. The location planning ontology contains eight main classes, each with subclasses. Figure 2 shows the details of the class relationships in the location planning ontology. We used more than twenty properties to model the ontology.

For this research, Web Ontology Language (OWL), recommended by the World Wide Web Consortium (W3C), is used to represent the ontology due to its capability of explicitly representing the concepts and their relationships. The travel ontology is modeled using Protg [8]. Protg is a free, open-source platform with a friendly user interface that provides a set of tools for constructing domain model and knowledge-based applications with ontologies.

3.3 Estimating users' preferences

When the first step of ontology building is performed, we can construct a Bayesian network for modelling users preferences. For these two reasons, a Bayesian network is used to estimate the preferred activities in the personalized recommendation system of attractions. We believe Bayesian network is a good method to support the combination of content-based and collaborative filtering approach.

When using BNs for knowledge model of users preferences, three steps exist for building a preliminary and usable Bayesian network: nodes selection, topology building and parameters setting. The first two steps can be addressed using rules predefined regarding the ontologies we have built, and the parameters elicitation can be learned from statistical data or estimated by a survey. We describe them in more detail in the following subsections.

3.3.1 Bayesian Networks

A Bayesian Network is a graph with arcs connecting nodes and no directed cycles (i.e., a directed acyclic graph), whose nodes represent random variables and whose arcs represent direct probabilistic dependencies among them [10].

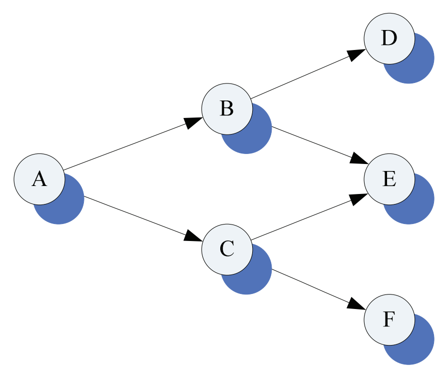

Fig. 3: An example Bayesian networks

Figure 3 shows a simple Bayesian network as an example. A directed arc between B and D denotes the fact that B influences the likelihood of D. Lack of directed arcs is also a way of expressing knowledge, notably assertions of (conditional) independence. For instance, lack of a directed arc between A and D encodes the knowledge that A does not influence D, only indirectly through the variable B. These causal assertions can be translated into statements of conditional independence: D is independent of A given B. In mathematical notation

$$P(d|b) = P(d|b,a) \tag{1}$$

Similarly, the absence of arc from B to C means that B will not influence C in a direct way.

These independence properties imply that:

$$P(a,b,c,d,e,f) = P(a)P(b|a)P(c|a)P(d|b)P(d|b)P(f|c)P(e|b,c) \tag{2}$$

i.e., that the joint probability distribution over the graph nodes can be factored into the product of the conditional probabilities of each node given its parents in the graph.

3.3.2 Qualitative stage

In this stage, the variables in the Bayesian network contain a set of factors influencing the preferred activities. In the travel domain, a great deal of literature can be investigated to identify the factors affecting preferred activity choice by theoretical and empirical study. In this paper, we have confined ourselves to a set of factors that are common mentioned by researchers in the travel domain, as described below. Due to the lack of space, we only introduce six main factors here: tour motivation, traveler type, age, occupation, personality and preferred activities.

After the relevant variables are defined, the relationships between the variables have to be established. In our developed simplified Bayesian network, four variables (age, occupation, personalities, and tour motivation) are root nodes without parents. Three variables, age, occupation, and personalities, influence the traveler type that, subsequently combine with tour motivation to influence the preferred activities. In other words, if the probability distribution of any of the above three variables (age, occupation, and tour motivation) is changed, the probability of the traveler type may change, and subsequently the probability distribution of the preferred activities is changed. If the probability distribution of the tour motivation is changed but others remain the same, the probability distribution of the preferred activities is also changed.

3.3.3 Quantitative and updating stage

First, in this stage, the probability distributions of the four root variables (age, occupation, personalities, and tour motivation) and the conditional probability tables for other two variables (traveler type and preferred activities) are assigned. For example, the conditional probability table for the variable traveler type is assigned based on the survey of travel behaviour. However, rather than doing a survey by ourselves, we summarized the empirical data by reviewing the existed research in the literature in travel domain, for example. The conditional probability table for the preferred activities is also assigned in the same way when the data for a user is limited at the beginning. When more and more recommendations are made for this user and feedback are obtained from that person, the conditional probability table can be reassigned based on that users travel behaviour. Second, the probability distributions of the two variables (traveler type and preferred activities) are calculated based on the results in the first step using Bayes theorem.

Then, the probability distribution of the preferred activities is updated using Bayesian theorem given the evidence of the four variables: age, occupation, personality, tour motivation. The updating stage involves two steps. First, the traveler type is updated given the first three variables (age, occupation and personality). Second, the preferred activity is updated given the last variable (tour motivation) with the updated traveler type. Thus, a users preferred activity is estimated based on that persons personal info (age, occupation, personality) and tour characteristics (tour motivation). In summary, this estimation combines factors from both users socio-demographic characteristics and tour characteristics. The recommendation is generated based on both the travel behaviour of the user and of other travelers. This provides a clear insight into the estimation of a users preferences. Each location destination is different. In order to recommend tourist attractions at a specific destination in the semantically enabled recommendation system for tour planning, an ideal way is to assign the conditional probability table destination-specifically, which is based on surveys in this destination.

Although we only build a part of the whole network, the rest can be constructed in a similar way.

3.4 Integrated Bayesian Knowledge Engineering Process

In order to build a high quality Bayesian network, a spiral engineering process is necessary as the knowledge acquisition cannot be performed immediately. We have developed an integrated Bayesian knowledge engineering process (I-BKEP) to address this problem. This process integrates knowledge elicitation from domain ontologies (travel in this paper) and automated knowledge discovery methods in a consistent way. I-BKEP methodology can be performed in a repeating cycle of design, development, operation and evaluation of BNs and the corresponding rules. Each evaluation phase is used to examine lessons learned and plan the next cycle of the development effort.

The proposed I-BKEP builds the target graphical model, which can be guaranteed by knowledge engineering process manager. The whole process includes six steps (see Figure 4).

In design phase the structure of the network is defined. This can be done using expert defined rules discussed above. Once a BN topology has been established, the next step is parameter estimation, involving specifying the conditional probability tables (CPTs) for each node. As it is usually to combine expert knowledge with existing data to set parameter, the expert must assess whether this kind of combination is acceptable. If not, it should go back to the combination process to make further decision. The design phase is supported by CIA and KE in our proposed framework.

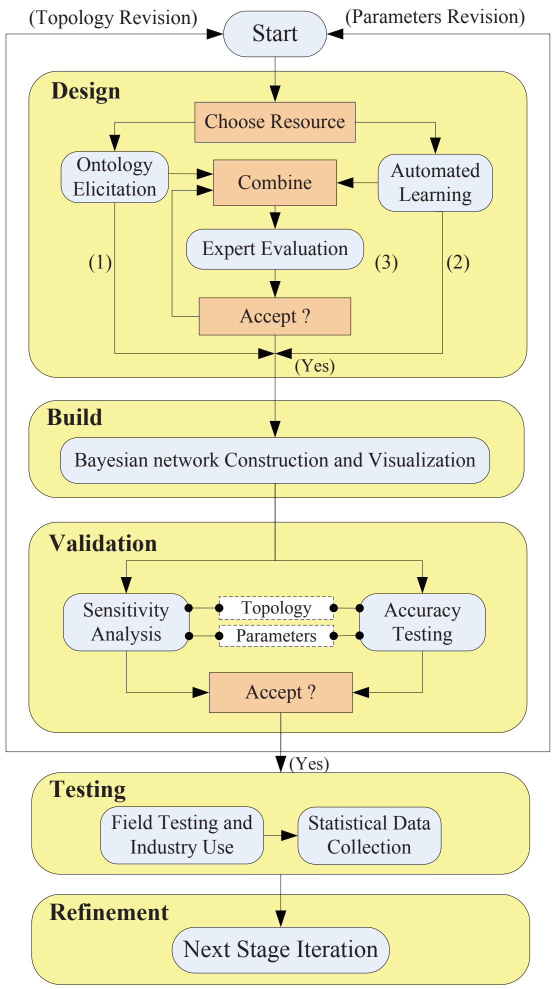

Fig. 4: Integrated Bayesian knowledge engineering process

Build step uses the results from the design phrase to build a Bayesian network with visualization by Protg [9]. If the established networks are valid (by testing in the next step), then they can be stored in the knowledge base for further use or evolvement. Validation step involves sensitivity analysis and accuracy testing. Sensitivity analysis can be critical in many commercial settings. This involves analyzing how sensitive the network is, in terms of changes in updated probabilities of some query nodes to changes in parameters and inputs. Measures for these can be computed automatically using BN tools, but these need to be evaluated by the expert in conjunction with the KE. Precision of measurement comes at a cost. We use entropy and mutual information as measure in our framework. Our algorithm computes and displays both the entropy of a specified query node and the ranked mutual information values for a specified set if concerned nodes in a BN, given a set of evidence for some observed nodes. Section 3.5 describes the detail of this sensitivity analysis method.

In testing phase, field testing of the Bayesian networks provide feedback and an opportunity to find and fix problems. Second, the product is used in industry and statistics are collected to allow for further refinement.

In application phase, the prototype network is used in industry and statistics are collected to allow for further refinement. Validation and testing must take place for the updated net is used.

In application refinement phase, the networks may be updated and refined to better fit its past and current data and correct for optimal decisions. The corresponding rules can be extended to cope with more general situations. Then, we go to the next stage in an iteration style. As the time goes on, the knowledge in the knowledge base becomes more and more accuracy, and consistent with the real system.

The proposed I-BKEP methodology can be regarded as a repeating cycle of design, development, operation and evaluation of Bayesian model. Each evaluation phase is used to examine lessons learned and plan the next cycle of the development effort. As network construction progresses, the expert's and knowledge engineer's understanding of the problem deepens and a shared language is created that facilitates communication. Exploring a prototype network's behavior on even a highly simplified problem fills in voids in the knowledge engineer's understanding of the domain and the expert's understanding of how a belief network thinks about the problem.

3.5 Sensitive Analysis Methods for I-BKEP

Since the probabilities elicited from domain engineer or history data are often systematically biased, it is interesting to investigate the reliability or robustness of Bayesian networks to see the effect of imprecision in probabilities on the network performance. In other words, model validation is necessary for a Bayesian network to achieve satisfying performance. For this purpose, a sensitivity analysis technique is often used to investigate the effect of probability parameter changes on the performance of Bayesian networks.

For testing and evaluating a built BN, even when adequate data is available, it is important to involve the domain engineer in evaluation. If engineer elicitation has been performed, a structured review of the probability elicitation is important. This procedure could involve: comparing elicited values with available statistics; comparing values across different engineers and seeking explanation for discrepancies; double-checking cases where probabilities are extreme (i.e., at or close to 0 or 1), or where the engineers have indicated a low confidence in the probabilities when originally elicited.

Sensitive analysis can be utilized to address the problem above [16]. It involves analyzing how sensitive the network is, in terms of changes in updated probabilities of some query nodes to changes in parameters and inputs.

Sensitivity analysis can be quantified using two types of measures, entropy and mutual information. Entropy, H, is commonly used to evaluate the uncertainty or randomness of a variable X characterized by a probability distribution, $P(x)$

$$H(X) = -\sum_{x \notin X} P(x) \log P(x) \tag{3}$$

Entropy measures assess the average information required in addition to the current knowledge to specify a particular alternative.

Another possible measure is the mutual information, which is closely related to the definition of entropy and conditional entropy. The conditional entropy of a pair discrete random variables Y given X is defined as

$$\begin{aligned} H(Y|X) &= \sum_{x} p(x) H(Y|X=x) \\ &= -\sum_{x} p(x) \sum_{y} p(y|x) \log p(y|x) \\ &= -\sum_{x} \sum_{y} p(x,y) \log p(y|x) \end{aligned} \tag{4}$$

It is easy to show that $H(Y|X) = 0$ when Y is a function of X, i.e., $Y = g(X)$. Since for all x with $p(x) > 0$, the value of Y is determined as $y = g(x)$ with probability $p(y|x) = 1$.

Mutual information measures the amount of information one random variable contains about another. For two random variables X and Y with a joint probability function $p(x,y)$, the mutual information $I(X;Y)$ is the relative entropy between the joint distribution $p(x,y)$ and the product distribution $p(x)p(y)$, and can be expressed by entropy and conditional entropy,

$$\begin{aligned} I(X;Y) = &\sum_{x} \sum_{y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ =&H(Y) - H(Y|X) \\ =&H(X) - H(X|Y) \end{aligned} \tag{5}$$

When Y is a function of X, $I(X;Y)$ degrades to the entropy of the responsive variable $H(Y)$, in other words, it becomes self information. In this case, the information gain of knowing X is zero to Y.

Using parameter sensitivity, we can identify which parameters in a Bayesian network are the most important with respect to the queries. Intuitively, when the different parameters undergo the same amount of variation, those with higher parameter sensitivity causes bigger changes in the query, and thus, affects the networks performance stronger. Efficient algorithms for sensitivity analysis in Bayesian networks made it possible to quickly identify the important probability parameters. The experiment results of the sensitivity analysis for the proposed Bayesian network will be described in detail in next section.

4 System Implementation and Scenarios

In this section, we describe the implementation of our location planning system and some scenarios of it.

4.1 System implementation

Our location planning system is implemented in Ruby on Rails (RoR). We leverage the RoR framework by using and extending various aspects. Using Rubys meta-programming support and the rich view capabilities of the RoR platform every recipe is converted into a Web application that can be customized to create rich AJAX Web applications. In addition, every recipe mashup can be exposed as a Web service (REST, RSS, or Atom).

The system architecture is implemented in a three-tiered way. The first tier runs on standard Web browsers, where the users send their requests, receive the recommended results, and send the feedback. The middle tier contains two types of servers: the application server and the Web server. In addition, the three key components for personalized recommendations, integrating heterogeneous information of tourist attraction, estimating travelers preference, and evaluating tourist attractions, are also included in this middle tier. The online information of tourist attraction is published by various travel information providers and is mashuped in the OWL based on the tourism ontology. The third tier contains spatial Web services such as Google Map, Yahoo Weather and WikiTravel.

We chose the Web Application Description Language (WADL), from Sun Microsystems, as a specification for describing REST services. A WADL document is designed to be a simple alternative to WSDL for use with XML/HTTP Web applications. It provides a description of Web applications in a simpler format than WSDL while also defining how to generate the URI for each operation and defining the format of the input and output parameters.

The prototype of this system has been implemented using Ruby on Rails and Google Map API for providing spatial data and GIS functions over the Web. In the prototype, an Apache web server is running as a Web server. The Netica [15] API Programmers Library, is embedded in the Web server side to estimate a travelers preferences in a Bayesian network. Integrated information about tourist attractions is represented in OWL.

4.2 Scenario

In the previous section, we built the location planning ontology which consists of a Bayesian network for inference. Now, we introduce the method of recommendation with ontology through scenario and show the query processing result in the interface. The

Table 1: Scenario of the travel requirement

Eric is living in New York and he wants to go to Tongji university by airplane today to attend a academic conference. We know that his preference are horse riding, golf, swimming in order from profile. However, it might be rainy in Shanghai. Please make a location planning for him.

Table 2: Scenario of the travel requirement

Language	OWL
Name	Mike
Location	Shanghai
Premise	Type sport LeisureSports
Query	HasRaining Sport?
Result	Indoor Swimming

Table 1 is a scenario to retrieve activities according to preference in our system.

Table 2 shows the recommendation result using OWL in travel ontology. We want to query about sports and it has been mapped to LeisureSports in premise. The query form in figure is following OWL grammar. As a result, we obtained IndoorSwimming. Because this preference sports like horse riding does not have value of HasRaining.

Therefore, we recommend IndoorSwiming refer to the third preference sports. As you see scenario, the rules and the relations between classes are important for recommendation.

Second, we show our Semantic travel search examples, and this search engine can answer the following questions.

(1) Please show me some interesting places in Beijing.
(2) Please show me some hottest attractions in Shanghai.
(3) Show me some interesting place I prefer to tomorrow.
(4) I want to go to Shanghai for a few days next week.
(5) I want to travel to China for a week.

Figure 5 is an example result of regarding question (5) queried in our tourism system. Besides recommendation for common items, such as cities, activities, we can recommend entire travel routing in our system, which is very useful for foreign traveler with little knowledge about the location.

4.3 Evaluation of the Model

The successful application of the BN for semantic mashup own to the Integrated Bayesian Knowledge Engineering Process described in section 3.5. In this section, we will show some experiment results of the evaluation of the BN during the engineering process.

4.3.1 Sensitive analysis

In this section, we first evaluate the BN using the sensitivity analysis methods described in section 3.6.

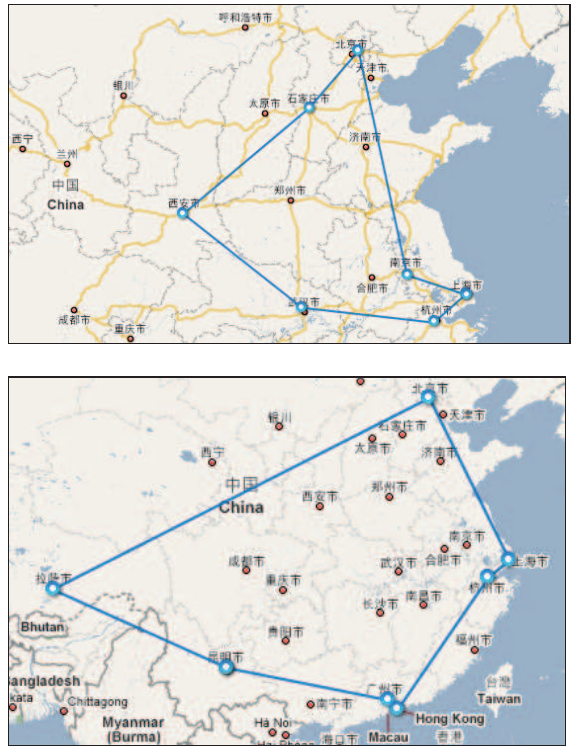

Fig. 5: Search results (these are two different proposed routs for the query: ”I want to travel to China for a week”)

Table 3: Sensitivity analysis for Bayesian network showing calculated entropy

Node Name	Entropy
Personality	0.156
Motivation	0.143
Age	0.098
Type	0.073
Occupation	0.052
Profession	0.038
Time available	0.027
Kind of trip	0.009
Gender	0.003
Money to spend	0.001
Accompanying person	0.001
Transportation	0.001

A typical output of sensitivity analysis, using entropy measures, is shown in Table 3. From the results we can evaluate and improve the model by engineers. Results from this table indicate that the *personality* and *motivation* are the variables having the greatest influence on the whole network, which both belong to operation system services.

In our China Tourism system, *personality*, *motivation*, *age* and *type* dominate the travel system. Results for mutual information discover the same results. This analysis results can be checked by domain expert (agreed) and utilized in the next iteration.

After the evaluation, the BN can be accepted for the next stage of development. This decision is not intended to be the end of the knowledge engineering or prototype development process.

4.3.2 Predictive accuracy

One aspect of knowledge engineering is evaluation, which guilds further iterations of BN development. When data is available, it can be used for evaluation. The most common method of evaluation is to determine the predictive accuracy (PA) of the BN both in travel line recommendation and interesting activity recommendation, which measure the frequency with which the model node state (with the highest probability) is observed to be the actual value.

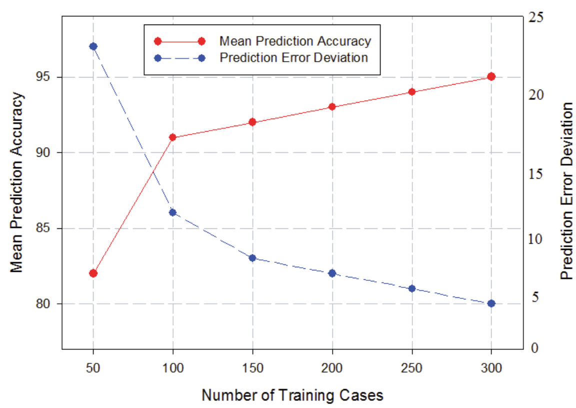

Fig. 6: Prediction accuracy of the Bayesian network

Our first result shows the effect of the number of cases on the construction of the Bayesian network model. To check the accuracy of the model, we performed a 15-fold cross validation. We observed the prediction accuracy of all the BN nodes, but present here only the results for the planning service. We started with 50 cases and went up to 300 cases. The criterion to achieve a suitable model was to get high prediction accuracy with low prediction error deviation. Figure 6 presents the results of this process and we found that 200 cases were sufficient to get a stable and accurate model, as increasing cases from 200 to 300 improves the prediction accuracy by less than 0.2%.

As the iteration of the knowledge engineering process been performed, the predictive accuracy of the BN become higher and higher. The reason behind this phenomenon is the knowledge becomes more and more accuracy regarding the varieties of data source and engineers experience. Figure 7 shows this process. The recommendation performance increased during the

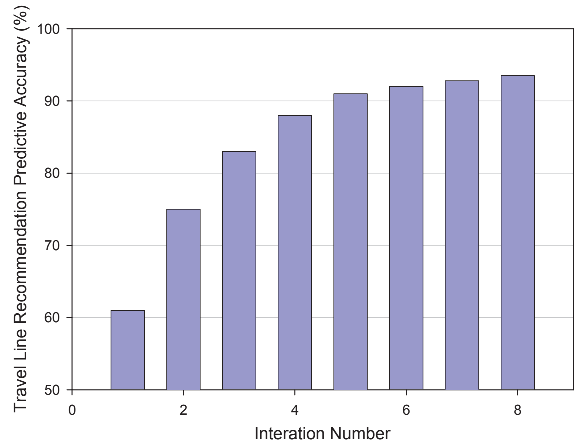

Fig. 7: Increase of the BN accuracy for travel line recommendation

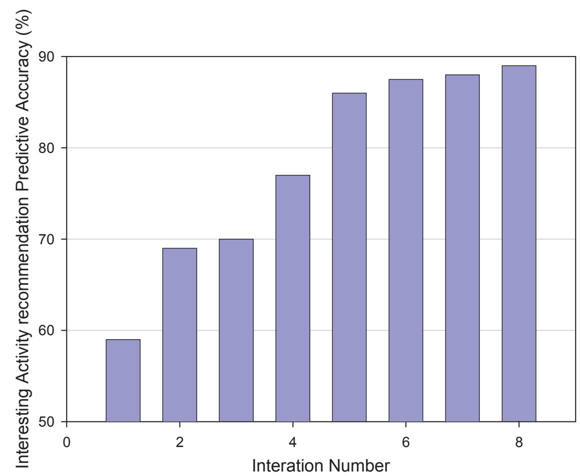

Fig. 8: Increase of the BN accuracy for interesting activity recommendation

iterations, which shows a promising way for travel router recommendation in our system. We have noticed that the first three iterations of refinement are very important which can significantly increase the performance of the BN, and after five iterations the performance of accuracy exceed 90%.

The interesting activity recommendation has the similar results which lead to a higher predict accuracy according to the check by engineers with the real case set (see Figure 8), which also shows the feasibility and practicability of our method.

5 Conclusions

In this paper, an ontology and Bayesian Network based methodology is proposed for location planning. And we present an intelligent recommendation system as an example to show the efficiency of the proposed method. The system offers the personalized recommendations of attractions at a given destination. Through the ontology of tourism, the system allows integration of heterogeneous online travel information. Based on Bayesian network technique method, the system recommends location attractions to a user by taking into account the behavior both of the user and of other users. Finally, using Web 2.0 based SOA paradigm, we develop a real China location planning system to show the usefulness of our method.

Acknowledgement

This work was supported by the National High-Tech Research & Development Program of China (Grant Nos. 2009AA012201), the National Natural Science Foundation of China (Grant Nos. 70903051, 61103068, 61174158, 61102059). The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

References

[1] S. Murugesan, Understanding Web 2.0, IT Professional, 2007, (7): 34-41.

[2] T. Berners-Lee, J. Hendler, O. Lassila, The Semantic Web, Scientific American, 2001, 284(5): 34-43.

[3] R. T. Fielding, R. N. Taylor, Principled Design of the Modern Web Architecture, ACM Transaction on Internet Technology, 2002, 2(2): 115-150.

[4] D. R. Fesenmaier, B. Pan, U. Gretzel, et al., Tourist decision model report, Project IST-2000-29474, 2002.

[5] E. Tomai, M. Spanaki, From Ontology Design to Ontology Implementation: A Web Tool for Building Geographic Ontologies. In Proceedings of AGILE 2005, Portugal, 2005, 281-290.

[6] L.Page, S.Brin, , et al., The PageRank Citation Ranking: Bringing Order to the Web. Stanford Digital Library Technologies Project, 1998.

[7] Y. Huang, L. Bian, A Bayesian Network and Analytic Hierarchy Process Based Personalized Recommendations for Tourist Attractions over the Internet, Expert Systems with Applications, 2009, 36, 933-943.

[8] Protg Ontology Editor (2004) http://protege.stanford.edu/

[9] F. Ricci, Travel Recommendation Systems. IEEE Intelligent Systems, 2002, 17(6): 55C57.

[10] K. B. Korb, A. E. Nicholson, Bayesian Artificial Intelligence. Chapman & Hall/CRC Press, 2003.

[11] H. David, M. Stefano, K. David, Piggy Bank: experience the semantic web inside your web browser, in: Proceedings of the International Semantic Web Conference (ISWC), 2005.

[12] H. Geng, Q. Gao, J. Pan, Extracting content for news web pages based on DOM, IJCSNS Int. J. Comput. Sci. Network Security2007, 7(2).

[13] J. Allsopp, Microformats: Empowering Your Markup for Web 2.0, Friends of ED, Berkeley, CA, 2007.

[14] W. Wang, H. Wang, B. Yang, et al., A Bayesian Knowledge Engineering Framework for Service Management, IEEE/IFIP Network Operations and Management Symposium (NOMS08), Salvador, Bahia, Brazil, April 7-11, 2008.

[15] GeNIe. http://genie.sis.pitt.edu/

[16] K. B. Laskey, S.M. Mahoney, Network Engineering for Agile Belief Network Models. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(4): 487-498.

[17] W. Wang, G. S. Zeng, D. Q. Zhang, et al., An Intelligent Ontology and Bayesian Network Based Semantic Mashup for Tourism, 2008 IEEE Congress on Services - Part I, 2008, 128-135.

[18] D.K.W. Chiu, Danny Kok, et al., Integrating Legacy Sites into Web Services with WebXcript, International Journal of Cooperative Information Systems, 14(1):25-44, 2005.

[19] D.K.W. Chiu, Yves T.F. et al., Towards Ubiquitous Tourist Service Coordination and Process Integration: a Collaborative Travel Agent System with Semantic Web Services. Information Systems Frontier, 2008.

[20] T.H.W. Lam, R.S.T. Lee, iJADE FreeWalker - An Intelligent Ontology Agent-based Tourist Guiding System, Computational Intelligence for Agent-based Systems, 2007, 103-125.

Daizhong Tang is the Lecturer of Tongji University. he is the deputy dean of Kingwai Real Estate Research Institute of Tongji University. Visiting scholar at Harvard University, Cambridge (USA), University of Illinois at Urbana-Champaign (USA). He received Ph.D degree in Management science and Engineering at Tongji University(China). His research interests are in the areas of applied mathematics and mathematical physics including the mathematical methods and models for information system development, Bayesian network. He has published research articles in reputed international journals of information and engineering sciences. He is referee and editor of several international journals.

Jiangang Shi is the professor of Tongji University. He received Ph.D degree in Management science and Engineering at Tongji University(China). His research interests are in the areas of Regional planning and development. He has published over 170 papers in related journals.

Wei Wang is the associate professor of Tongji University. Visiting scholar at Erlangen Nuernberg University (German), Aalto University (Finland), University of Wisconsin-Madison(USA). His research interests are in the areas of Data intensive cloud computing, Computer and Information Security and Parallel and heterogeneous architecture. He has published research articles in reputed international journals of IEEE Transaction series. He is referee and editor of several international journals.