A Probabilistic Knowledge-Based Information System for Environmental Policy Modeling and Decision Making

Hamid Jahankhani*, Elias Pimenidis, and Amin Hosseinian-Far

School of Architecture, Computing & Engineering, Docklands Campus,
University of East London- London E16 2RD, UK
{Amin,E.Pimenidis,Hamid.Jahankhani}@uel.ac.uk

Abstract. Decision making for setting new policies is a challenging process as the current policy making system is utterly flawed. A policy is introduced by the decision maker when the problem domain was fully consulted by experts in the field. Not always all the consultants and advisers agree on details or even basics of such a course of action. The need for an intelligent predictive system is emerging. Policy making on environmental issues are even shoddier as the environmental systems are habitually complex, and adaptive; and introduction of new technologies can easily affect the guiding strategies already taken. This paper outlines the principles of Knowledge Management Systems. It then reflects on Influence Diagrams' suitability for construction of such an information system through the use of the London Plan case study. An application of such a system is outlined by means of a probabilistic knowledge based IS which is developed by Influence Diagrams and can be utilized as an Environmental policy modeler and/or DSS.

Keywords: Knowledge Management System, Decision Support System, London Plan, Influence Diagram, Tacit and Explicit Knowledge.

1 Introduction to KMS

There are various applications for Knowledge Management Systems (KMS) theories including but not limited to Distributed Databases, Ontology, and Artificial Intelligence (AI). Knowledge Based systems which are one of the software tools in Intelligent Systems are the instruments to manage the knowledge [1]. The main focus would be on the Artificial Intelligence in this research.

Knowledge Management Systems can be classified into two main categories; Distributive and Collaborative aplications. Distributive knowledge management system is where database plays a vital role in shaping the knowledge base, hence there is a structured formal database involve. On the other hand the collborative knowledge management system partakes an informal internal knowledge[2].

2 Types of Knowledge and Knowledge Management Systems

Some experts still argue on the defintion of knowledge itself. knowledge is summation of information, skills, experience, and personal capabilities [3]. Other

* Corresponding author.

L. Iliadis et al. (Eds.): AIAI 2012, IFIP AICT 381, pp. 136–145, 2012.

scholars also attempted to categorize and formulated knowledge and its types. Another categorization of knowledge is outlined in Fig. 1:

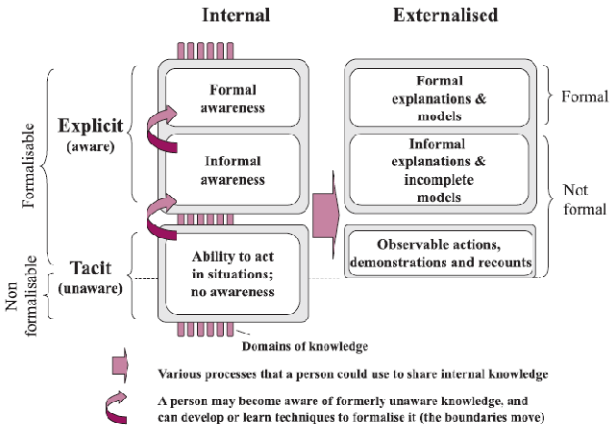

Fig. 1. Categories of Knowledge [4]

Defining knowledge is not enough as the knowledge should be retained and gathered. There are various frameworks for knowledge acquisition. The common knowledge acquisition framework is illustrated in Fig. 2:

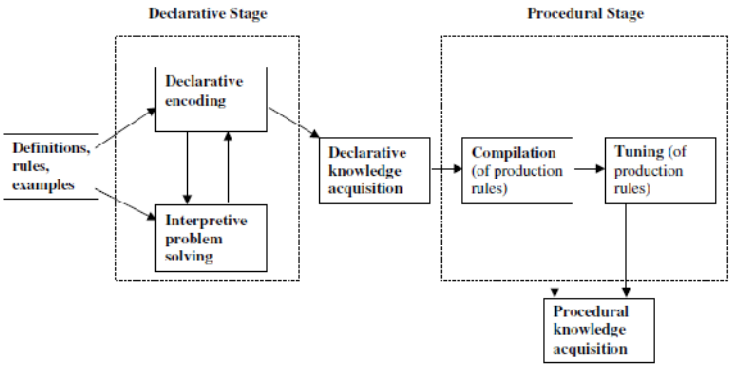

Fig. 2. Knowledge Acquisition Framework and Stages [5]

According to this general framework, the first step is the declarative stage where the rules, standards, examples and definitions should be clearly defined. Once the standards are declared, the problem solving and encoding stage begins, still within the declarative stage. Up to this point the knowledge is still in the declarative knowledge acquisition mode. In the start of the procedural stage, compilation should be performed on the declarative acquitted knowledge. Once the errors are fixed and the compilation is tuned, then the procedural knowledge acquisition is completed [5].

Furthermore, knowledge can be transformed while in use. The general categorization for knowledge transformation is internal, external and socialized knowledge (Fig. 2). The knowledge is initially in the internal form in tacit format. The tacit format indicates that the knowledge still contains areas where the use is not fully identified. The tacit knowledge within the internal domain will be transformed into explicit knowledge where the standards and usages will be clarified. By having the formal explanations and models the knowledge can be externalized. The externalization on its own does not have to be formal at all circumstances. Informal explanations and observations in the external domain based on the explicit internal knowledge can still provide a means for formalization. The set formal external knowledge which is explicit can be widely used and become a socialised knowledge format [6].

A dynamic system model which has the learning characteristic, complexity elaboration and non-linear features will be able to correspond to the research questions within the problem domain. The complexity, non-linearity, and adaptive characteristics are the main factors that should be considered for selection of proper tool. Fig. 3 represents different software tools for Artificial Intelligence. As it is preferred to use a predictive design one of the following should be selected for the design. Generally aritificial intelligent systems are divided into two main categories which are Knowledge Based systems and Computational Intelligence systems.

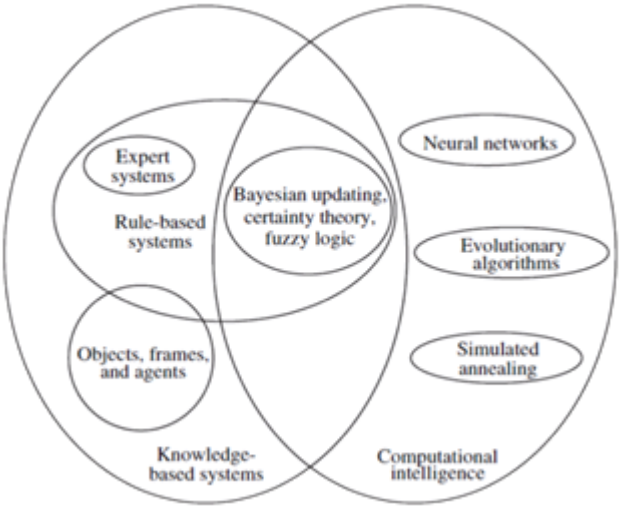

Fig. 3. Artificial Intelligence areas [7]

Some characteristics of the sustainability can be modeled using other types of system modeling techniques; for instance using cybernetic system approach through state space method for socio- ecological system. The latter has been modeled for the climate change by Schellnhuber [8].

3 Problem Area

Policitcians are having a demanding era in facing the sustainability challenges. Policy making as a rule happens after long periods of consultancies. Not always, all the consultants and polictical advisers come to a common ground in different scenrios. The use of predictive information Systems can be valuable in assisting the decision maker. These ISs have be predictive in a way that can approximate the outcomes of different decisions in a given scenrio. Modeling environmental states is quite a challenging process. The tradeoff between complexity, efficiency and accuracy of the Information System used for development of a DSS is crucial. Use of probabilistic inference is often overlooked. Influence Diagrams practices as the primary means for developing the knowledge base for complex environmental scenarios should be tested. Appraisal of Influence Diagrams and ID evaluation techniques can be tested by means of a case study research design type. The London Plan as a case study has had various discussions since 2004[9]. Using microeconometric theories, the required investment for policy 4A.2 has been calculated between 40 to 80 £Billion. Since the case study has these estimates using other quantitative modeling techniques, using Influence Diagrams and probablistic inference can be validated for these kind of modelings. Bayesian networks can be utilised for development of such a knowledge base, but chance nodes are required for development of a DSS. Therefore, Influence Diagram is a fine choice for the knowledge base construction. The objectives of this research would be:

1. Knowledge management representation of the knowledge body behind the London Plan sustainability domain. The London Plan is the case study used for validation of the proposed model.
2. Identifying an appropriate quantitative approach for modeling the financial domain of The London Plan.

4 Probabilistic Networks for Knowledge Types

Real-life state of affairs mostly modelled as group of entities demonstrating random variables in a “probabilistic network”. Clever graphical illustration of dependence and independence relations between accidental variables is a “probabilistic network”. Area of random variables could, for instance, help decision makers to identify the most beneficial decision in a given situation from the basis of a decision support system.

Probabilistic networks processes and symbolize probabilistic data. Representative elements of a probabilistic network are a quantitative and a qualitative element. The qualitative component sets (conditional) belief and independence assertions along with a set of chance variables, informational fondness, and preferred relations [10]. Graphical language visually encoded the statements of (conditional) dependence and independence, information preference, and favourite relations. On the other hand, the quantitative component identifies the potency of dependence relations by means of probability and utility theories [11].

The illustrative depiction of a probabilistic network, explains knowledge of a problem area in a clear-cut manner [12]. The illustrative depiction is perceptive and easy to understand, making it an ideal tool for passing statement of domain knowledge among experts, users, and systems. Therefore, the formalism of

probabilistic networks is becoming an ever trendier domain knowledge representation for interpretation and decision making under uncertainty [11]. In this model, tacit and explicit knowledge are extracted using narrative extraction from plan policy reports and the discussions. The decision node is derived from the policy objectives. Therefore in this scenario, 60% CO2 reduction as the aim objective makes the decision node. The suitability of Influence Diagram is assessed as follows.

5 Suitability of Influence Diagrams for Environmental Policy Modeling

Despite having various modeling techniques in the field of sustainability, probabilistic inference and Bayesian networks models should be given high priority. The complexity of sustainability scenarios and understanding the systems' resilience and boundary paradox are essential when analyzing and designing the case studies [12]. Among the quantitative techniques, agents and multi agent systems are quite useful as they can bring together the components of the system using agents' characteristics. Furthermore the artificial intelligence techniques used in multi agent systems would enable the system to adapt itself with the changes faced from outside the boundary. But there is a major drawback with multi-agent systems as they are too expensive for building large Decision Support System applications. Although they can fairly fit fine in a CAS (Complex Adaptive System) methodology in theory [13], but when it comes to practice, they are not responsive to all of the components. Use of neural networks as the tool for modeling the inputs and the outputs based on a network of linked components sound very functional. But the major drawback with the neural networks is when it comes to its training. Training of the network in a sustainability scenario is almost impossible. The longitudinal study may help for the training, but long intervals and wait time is required for the testing. Hence practically, they are not usable in large complex systems where training is nearly impossible. The system's dynamic models are also other quantitative techniques for sustainability modeling. The difficult nature of developing a system using systems' dynamic, has made it again impossible for the developer to develop the model. System dynamic might be useful for modeling some small trade-offs in the environment where increase in one component would lead to a drop in another. Stella might be a good IDE for these small projects; but again, if the system becomes large with various elements in the knowledge base, analysis, design and development of the model would be unattainable [14].

On the other hand, probabilistic inference, e.g. influence diagrams are fine tools for modeling sustainability. The adaptively of the model is an issue with probabilistic networks as the base for the model, but that can be easily fixed by incorporating an agent based engine. The use of goal oriented knowledge management systems theories would be a good further work for this study. This research study by means of the London Plan case study validates that despite having a complex scenario; modeling with ID would be feasible. The model would be a formal model where replication, testing and validation are possible using the available algorithms. The predictive nature of the Influence Diagrams, very good graphical interface for non-expert users, and also accurate mathematical and probability layer would be one of the finest approaches to model sustainability systems with complex nature. It still fit into the Complex Adaptive

System structure if the agents are incorporated to give the adaptively feature to the system. The use of knowledge management theories would help to build the knowledge base for the model [15]. Selection of a specific probabilistic network on its own should be rationale. The reason for selection of Influence Diagrams over Binominal Trees and Decision trees is explained using more details on the other two techniques; although they can be converted to each other in many scenarios.

6 KMS for the London Plan Case Study

The final model can be implemented as decision support system for the stakeholders. It also has the capability of being disseminated as a web based DSS tool via web services techniques. There is an overlap is between probabilistic networks and fuzzy systems. The ID uses the same theories, but incorporates the set graphical interface for better understanding and modeling the problem domain. Financial analysis of the London Plan policy 4A.x (X=1, 2, 3) is an appropriate case study for this validation, although other case studies might be used. The reason for selection of such a case study lies within the gap in discussions and analysis of the GLA plan. The London Plan Policy 4A.X proposes 60% reduction in London CO2 compared to 1990 base, using CHP (Combined Heat & Power), Energy Efficiency, and Renewable Energy Systems expending solar and wind [9]. The data extraction is done using a theoretical narrative extraction, and the design base of this research would be quantitative practical base. Fig. 4 outlines the high level ID for this scenario.

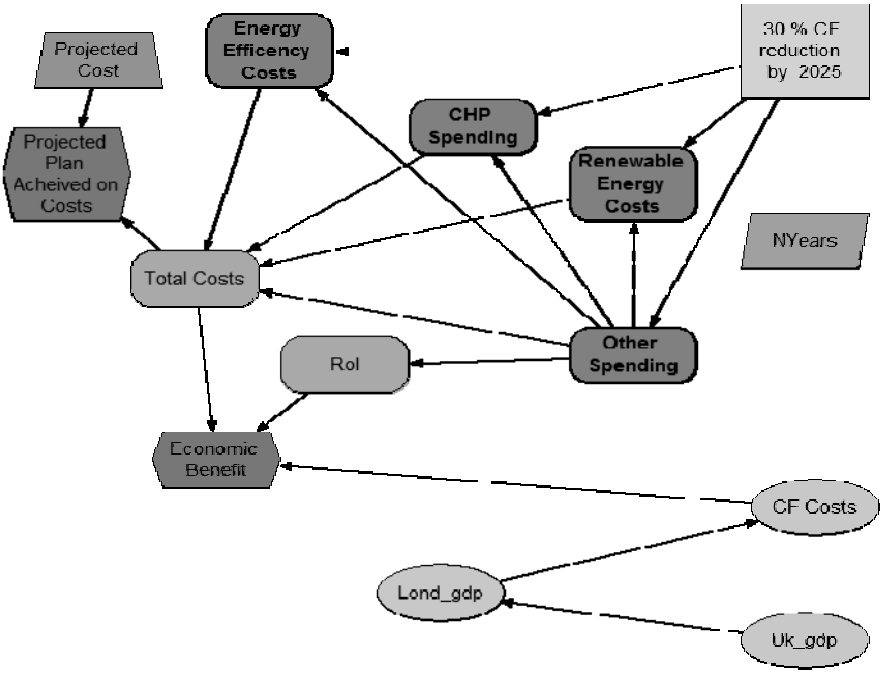

Fig. 4. High Level Influence Diagram as the Knowledge Base for Policy 4A.x Financial Assessment

Reasoning in such a model is dependent on the type of scenario under analysis. In this scenario, there is no 'Explaining Away' reasoning case between the nodes. Although, there are converging connections in the node orders, but explaining away cannot be inferred in the Extended ID developed. The converging inference can be assessed when the node orders are reversed using the reversal technique. There is no successor for this decision node and therefore the decision node removal cannot be considered for evaluating this ID. The reversal and node removal including barren node removal for this influence diagram can be further works of this piece of research. Casual reasoning in the model is where the UK-GDP has statistical dependence on the London GDP and that continues to all other proceeded nodes. There is various inductive reasoning in the ID where the parent nodes are not dependent to each other, i.e. 'Total Costs'.

$\varepsilon = \hat{A}^{\alpha}.\hat{C}^{1-\alpha}$ is the efficiency function for an ID. The parameters involved are Accuracy (A), and Complexity (C) [16]. In addition, α is the value set by the decision maker. This formula should be looked at with caution, as the complexity and accuracy of the ID should be normalized first [17]. Cobb's formula concerning the accuracy and complexity functions are as follows:

$$\hat{A} = N_{min}\frac{(N_{max}-N_{min}).(A-\underline{A})}{\overline{A}-\underline{A}}$$

$$\hat{C} = N_{min}\frac{(N_{max}-N_{min}).(C-\underline{C})}{\overline{C}-\underline{C}}$$

A represents accuracy and C represents complexity. The consideration is that when the accuracy is maximized, the complexity of minimized and vice versa. N_{min} represents 1 and N_{max} represents 2 in Cobb's research. The Accuracy and Complexity is scaled between a minimum and a maximum in order to assess a trade-off. The complexity of the ID expressions can be calculated using LeafCount function of the Mathematica IDE. Mathematics is software developed by Wolfram which can evaluate the complexity of functions and graphs. The calculation is simple and is counting the variables of any approximation within the ID. It does not only count the variables, but also the expressions defining the function. Therefore LeafCount is a function which counts the number of words, variables and constants in an expression [17]. The rationale behind this calculation is consideration of the memory required to process the function. The Complexity of an ID then is the summation of all individual complexities:

$$C = \sum_{i=0}^{n} C_i = \sum_{i=0}^{n}\sum_{j=0}^{m_i} \mathcal{L}\{\theta_{ij}\}$$

Using the LeafCount function we can simply determine the complexity of expressions and sub-expressions of the London Plan Policy 4A.x. For instance:

```
In[1]:=LeafCount [399943*Num_london_propertie]
Out[1]=3
In[2]:=Level [399943 + * + Num_london_propertie, Heads → True]
Out[2]= {Multiply, 399943, Num_london_propertie}
```

The complexity of the RoI node expression is 3. That is summed up to complexity outputs from the LeafCount function in Mathematica on other nodes. If there are

combinatory expressions in the ID, then the NestList function of Mathematica, lists all the expression in combinations and then the LeafCount can be run. The complexity of the whole Extended Influence Diagram for considering the Leafcount function is done manually and the number 50 is reached.

For finding Accuracy of Influence Diagram accuracy, the mean squared error between the analytical decision rule and the ID decision rule should be calculated.

The found Accuracy and Complexity values are then traded off using the stated formulas. Cobb's minimum and Maximum (N) and be scaled using other values, but the technique remains the same. The efficiency of an ID can also be affected by the decision makers' amendments using α variable [17].

This model also validates the approximations on predicting the financial costs of the policy which gives practical values to this research study. The result of the research indicates the estimated costs of the London plan policy concerning CO2 reduction has been within the discussed outcomes. The output of the simulation states that the London plan would cost £Billion 68.84. This number is derived after subtraction of £Billion 12.06 Billion pounds return on investment from the total costs which was £Billion 80.9. The simulations are performed within Analytica IDE, but other environments can also be utilized for simulation.

However, Analytica is a simulation-based tool, and currently cannot prescribe an optimal solution for the detection network. Although unavailable in software tools, influence diagrams with continuous variables can be found in the literature, Again, none of the works is applicable to the parallel detection network structure. Now consider a variation where we remove the continuous variables, i.e., the measurements from the sensor, and transform the local decision makers to be chance nodes.

This yields an alternative model, which is equivalent a similar structure was introduced by Heckerman to analyze the value of information for diagnosis. In this configuration, the subordinate decision makers provide evidence or local decisions [18].

7 Conclusion, Limitation and Further Works

There are number of limitations to this model. One the limitations of this research study would be the all-embracing complexities involved. The complexities do not only arise from the financial domains of the policy, but also the social and environmental facets as well.

Another constraint of this research project has been the consideration of only the financial facets of the policy plan. The social limitations and all environmental restraints reside outside the boundary of this study. Knowing that this project involves an open system analysis, therefore, not considering those two facts would restrain the overall result of the research work.

On the case study, there are some zones where the plan policy has not clarified fully. For instance the choice of 60% reduction by 2050 within policy has not been explained. In addition to that a comparative financial analysis has not been provided

in the policy. This could be used as a framework for comparing the results of this project and other similar projects and the policy plan results.

There are some limitations in Knowledge Management Systems.These limitations in KMS are defined as:

1. Tacit knowledge cannot be easily monitored and managed,
2. Involvement of the stakeholders in a dynamic and up to date management of their knowledge,
3. Interactions between stakeholders might be limited,
4. System is not necessarily adaptive [20].

Recent research on goal oriented models suggests that Knowledge Management Systems can be developed in a goal oriented friendly format. The initial proposal for goal oriented KMS was introduced in 2004, but still various research is taking place concerting this topic. In theory the goal oriented KMS would have the adaptively, innovation and replication characteristics [21]. Although this is out of scope of this study, but a future work on ID model developed in this research work might be implementation of its goal oriented KMS which has the mentioned here characteristics.

References

1. Valente, G.: Artificial Intelligent Methods in Operational Knowledge Management, Torino (2004)
2. Zack, M., Serino, M.: Knowledge management and Collaboration Technologies. The Lotus Institute, Lotus Development Corporation (1996)
3. Baker, M., Baker, M., Thorne, J., Dutnell, M.: Leveraging human capital. Journal of Knowledge management 1(1), 63–74 (1997)
4. Kalpic, B., Bernus, P.: Buesiness Process Modeling through Knowledge Management Perspective. Journal of Knowledge Management 10(03), 40–56 (2006)
5. McCall, H., Arnold, V., Sutton, S.: Use of knowledge management Systems and the Impact on the Acquisition of Explicit Knowledge. Journal of Information Systems 22(02), 77–101 (2008)
6. Kernstock, P.G.: A Web-Based knowledge Management System for information technology Education. Texas State university, Texas (2006)
7. Hopgood, A.: Intelligent Systems for Engineers and Scientists, 2nd edn. CRC press, New York (2001)
8. Schellnhuber, H.J.: Earth system' analysis and the second Copernican revolution. Nature (1999)
9. Greater London Authority: The London Plan (2011), from London. Gov., http://www.london.gov.uk/priorities/planning/vision/london-plan/replacement-process (retrieved 2011)
10. Uffe, B., Kjaerulff, A.L.: Bayesian Networks and influence Diagrams: a guide to construction and analysis. Springer (2007)
11. Lerz, J.L.: The Case for Using Probabilistic Knowledge in a Computer Chess Program, from Verizon.net (2012), http://mysite.verizon.net/vzesz4a6/current/id309.html (retrieved February 2011)

12. Hosseinian-Far, A., Jahankhani, H., Pimenidis, E., Wijeyesekera, D.C.: Reflections on Modeling Systems' Resilience. In: 5th SASTech Intl. Symposium, Mashhad, Iran (2011)
13. Hosseinian-Far, A., Jahankhani, H., Pimenidis, E., Wijeyesekera, D.C.: Reflections on the Need for an Improved Quantitative Approach. In: Advances in Computing and Technology Conference, London (2011)
14. Hosseinian-Far, A.: A Systemic Approach to an Enhanced Model for Sustainability. PhD Thesis (2012)
15. Hosseinian-Far, A., Pimenidis, E., Jahankhani, H., Wijeyesekera, D.C.: Financial Assessment of London Plan Policy 4A.2 by Probabilistic Inference and Influence Diagrams. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H. (eds.) EANN/AIAI, Part II. IFIP AICT, vol. 364, pp. 51–60. Springer, Heidelberg (2011)
16. Cobb, B.R.: Measuring Efficiency in Influence Diagram Models (2008)
17. Wolfram: The Mathematica Book. Wolfram (2003)
18. Baye, M.R.: Managerial Economics and Business Strategy. Mc-Graw Hill, New York (2006)
19. Heckerman, D., Shachter, R.: A Decision Based View of Casualty. Microsoft Research (1995)
20. Nabeth, T., Angehrn, A.A., Roda, C.: Enhancing Knowledge Management Systems with Cognitive Agents. Journal of information Systems in Management 8(2) (2003)
21. Gray, P.H., Meister, D.B.: Knowledge sourcing effectiveness. Journal of Management Science 50(06), 821–834 (2004)