THE BAYESIAN WEB

KENNETH BACLAWSKI

College of Computer and Information Science
Northeastern University
360 Huntington Avenue
Boston, MA 02453 USA
E-mail: Ken@Baclawski.com

The Semantic Web is an extension of the current World Wide Web in which information is given a well-defined meaning, so that computers and people may more easily work in cooperation. This is done by introducing a formal logical layer to the Web in which one can perform rigorous logical inference. However, the Semantic Web does not include a mechanism for empirical, scientific reasoning which is based on probabilistic inference. Bayesian networks are a popular mechanism for modeling uncertainty and performing probabilistic inference in biomedical situations. They are a fundamental probabilistic representation mechanism that subsumes a great variety of other probabilistic modeling methods, such as hidden Markov models and stochastic dynamic systems. In this paper we propose an extension to the Semantic Web which we call the Bayesian Web that supports Bayesian networks and that integrates probabilistic inference with logical inference. Within the Bayesian Web, one can perform both logical inference and probabilistic inference as well as reconcile stochastic models and perform statistical decisions. We discuss how the Bayesian Web would be used for representing and reasoning within biomedical ontologies.

1. Introduction

Probabilistic modeling has a long history, and it is the basis for the empirical methodology that has been used with great success by modern scientific disciplines. Stochastic models have traditionally been expressed using mathematical notation that was developed long before computers and graphical user interfaces became commonly available. A Bayesian network (BN)[9] is a graphical mechanism for specifying the joint probability distribution of a set of random variables. As such BNs are a fundamental probabilistic representation mechanism for stochastic models. The use of graphs provides an intuitive and visually appealing interface whereby humans can express complex stochastic models. This graphical structure also has been used in the design of efficient algorithms for data mining, learning and stochastic inference.

The range of potential applicability of BNs is large, and their popularity has been growing rapidly. BNs have been especially popular in biomedical applications where they have been used for diagnosing diseases[5] and studying complex cellular networks[3], among many other applications. The BNs that have been developed for disease diagnosis are especially large.

The Semantic Web (SW) was proposed by Tim Berners-Lee and his colleagues[2] as a means of introducing formal semantics to the World Wide Web. One of the fundamental features of the Web is its support for resource identifiers (URIs) which make it possible for documents to refer to each other as well as for multiple documents to make references to the same resource. The SW goes one step further and adds formal semantics to the resources identified by URIs and to the links between resources. All reasoning in the SW is formal and rigorous.

Although very large BNs are now being developed, each BN is constructed in isolation. Interoperability of BNs is possible only if there is a framework for one to identify common variables. In this paper we propose to use the SW as the basis for supporting BN interoperability. This is done by adding BN layer to the SW. We call the resulting framework the Bayesian Web (BW). This framework makes it possible to perform operations such as:

- Use a BN developed by some other group almost as easily as one now navigates from one Web page to another.
- Make stochastic inference and statistical decisions using information from one source and a BN from another source.
- Fuse BNs obtained from disparate sources by identifying variables that measure the same phenomenon.
- Reconcile and validate BNs by checking mutual consistency.

This paper begins with some background material on BNs and stochastic inference including some examples from medical diagnosis. In Section 3 we discuss the basic requirements for interoperability of BNs which are the motivation for this paper. Section 4 then gives some background on the SW. In Section 5 we give a concrete proposal for a BW which combines BNs with the SW. The paper ends with some conclusions and future directions for this work.

2. Bayesian Networks and Inference

A BN is a graphical formalism for specifying a stochastic model. The random variables of the stochastic model are represented as nodes of a graph. We will use the terms "node" and "random variable" interchangeably. While one would think that the notion of a random variable is unambiguous, in fact it is a combination of two different concepts. First, there is the phenomenon that is being observed or measured, such as one toss of a coin or the measurement of a person's blood pressure. The second concept is the probability distribution of the phenomenon. It is the combination of these two notions which is the mathematical concept of a random variable. The relationship between the phenomenon and its probability distribution is many-to-many. Many phenomena have the same probability distribution, and the same phenomenon can be distribution in many ways. The reason why a phenomenon does not uniquely determine its probability distribution is due to the notion of *conditioning*. As one observes related events, the distribution of a phenomenon changes. The phenomenon is the same, what changes is the knowledge about it (or more precisely about one instance of it).

The edges denote dependencies between the random variables. This is done by specifying a *conditional probability distribution* (CPD) of a node by specifying the conditional probability of each value of the node given each combination of values of the nodes at the other ends of the incoming edges. The nodes at the other ends of the incoming edges are called the *parent* nodes. A CPD is a function from all the possible values of the parent nodes to probability distributions on the node. Such a function has been called a *stochastic function*[6]. If a node has no incoming edges, then its CPD is just the probability distribution of the node. It is also required that the edges of a BN never form a directed cycle: a BN is *acyclic*. If two nodes are not linked by an edge, then they are independent.

Some of the earliest work on BNs, and one of the motivations for the notion, was to add probabilities to expert systems used for medical diagnosis. The Quick Medical Reference Decision Theoretic (QMR-DT) project[5] is building a very large (448 nodes and 908 edges) BN. Consider, for example, the BN shown in Figure 1. The BN is a very small diagnostic BN which specifies a stochastic model with four random variables: (1) Flu, i.e., a patient has influenza, (2) Cold, i.e., a patient has one of a number of milder respiratory infections, (3) Perceives Fever, i.e., the patient perceives having a fever, (4) Temperature, the continuous random variable represent-

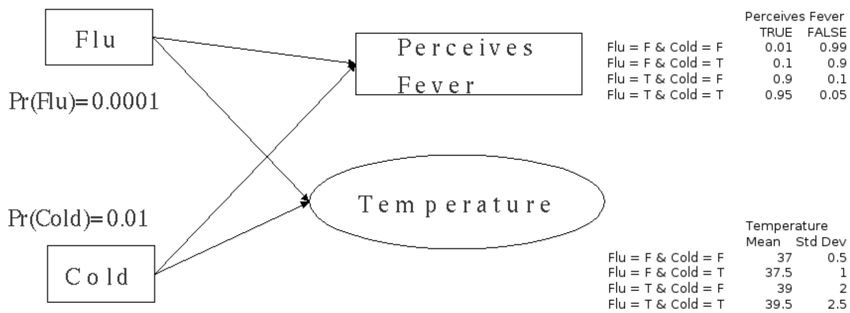

Figure 1. Example of a BN for medical diagnosis. Rectangles represent discrete random variables and ovals represent continuous random variables.

ing a measurement of the patient's body temperature. Note that three of the random variables are Boolean, the simplest kind of discrete random variable, and that the fourth random variable is continuous. Two of the nodes have no incoming edges, so their CPDs are just PDs, and because the nodes are Boolean, they can be specified with just one probability. We assume that $Pr(Flu) = 0.0001$ and that $Pr(Cold) = 0.01$, reflecting the fact that influenza is far less common than the common cold.

The CPD for the Perceives Fever (PF) node has two incoming edges, so its CPD is a table that gives a conditional probability for every combination of inputs and outputs. The CPD for the Temperature (T) node has two incoming edges, so its CPD will have 4 entries as in the case above, but each entry is a continuous probability distribution.

BNs have a number of other names. One of these, *belief networks*, happens to have the same acronym. BNs are also called probabilistic networks, directed graphical models, causal networks and "generative" models. The last two of these names arise from the fact that the edges can be interpreted as specifying how causes generate effects. One of the motivations for introducing BNs was to give a solid mathematical foundation for the notion of causality. In particular, the concern was to distinguish causality from correlation. A number of books have appeared that deal with these issues such as one by Pearl[10] who originated the notion of BNs. For causation in biology see[11]. Other books that deal with this subject are[4,13].

One of the main uses of a BN is to make deductions. A BN acts something like a rule engine. In a rule engine, one specifies a collection of if-then rules, called the rule base. One can then input a collection of known facts (typically obtained by some kind of measurement or observation). The rule engine then explicitly (as in a forward chaining rule engine) or implicitly (as in a backward chaining rule engine) infers other facts using the rules. The set of specified and inferred facts form the knowledge base. One can then query the knowledge base concerning whether a particular fact or set of facts has been inferred.

As in a rule engine, one can specify known facts to a BN (via measurement or observation), and then query the BN to determine inferred facts. Specifying known facts is done by giving the values of some of the random variables. The nodes that have been given values are termed the *evidence*. One can then choose one or more of the other nodes as the *query* nodes. The answer to the query is the JPD of the query nodes given the evidence. Since a BN is a mechanism for representing a JPD, the result of a BN inference is BN on a subset of the nodes of the original BN.

3. Requirements for Bayesian Network Interoperability

The most fundamental requirement of BN interoperability is to have a common interchange format. However, this alone would not be enough for one to automatically combine data and BNs from different sources. In this section we discuss the requirements for BNs to be fully interoperable in the sense discussed in the introduction.

The following are the requirements for BN interoperability and the proposed BW:

(1) Interchange format. There already exists an format for representing BNs, called the XML Belief Network format (XBN)[14]. This XML file format was developed by Microsoft's Decision Theory and Adaptive Systems Group. This format evolved from a standardization effort to develop the Bayesian Network Interchange Format (BNIF).
(2) Common variables. It should be possible for the same variable to appear in different BNs. For example, whether a person has the flu should be the same variable no matter which BN it appears in. Being able to specify or to deduce that two entities are the same is a fundamental feature of the Semantic Web. Of course the context within which a BN is valid affects the meaning of the variable. For example, one might be interested only in the occurrence of the flu in Spain in 1918. This would be very different from the flu in Australia in 2004.
(3) Annotation and reference makes it possible to specify the context

of a BN. In so doing one also specifies the meaning of the variables. One should be able to refer to a BN and for a BN to refer to other information. In other words, the BN is itself an entity about which one can make statements. Annotations are also important for authentication and trust. The BN itself can claim that it arises from a source that one trusts, but one would only believe it if a trusted source refers to the BN.

(4) Open hierarchy of distribution types. New probability distributions and conditional probability distributions can be introduced by subclassing other distributions.

(5) BN components. A BN can be constructed from known pieces. It can also be constructed by instantiating a template (possibly more than once). A BN component is a partially specified BN.

(6) Information fusion. Multiple BNs can be combined to form new BNs. This is a very different form of combination than component-based construction. This technique is called information fusion. Inference is, in fact, a form of information fusion because the output of inference is a JPD on the query nodes which can be expressed as another BN.

4. The Semantic Web

The increasing diversity and complexity of information available electronically has spurred interest in the notion of formal ontologies and in automating many ontology-related activities that were traditionally performed manually. Web-enabled agents represent one technology for addressing this need[8]. These agents can reason about knowledge and can dynamically integrate services at run-time. Formal ontologies are the basis for such agents.

The Resource Description Framework (RDF)[7] and the Web Ontology Language (OWL)[12] are ontology language standards developed under the auspices of the World Wide Web Consortium. RDF is the basic language with the minimum number of constructs necessary for expressing ontologies. OWL adds features to RDF in a series of three versions (or levels), called OWL Lite, OWL-DL and OWL Full.

The DL in OWL-DL stands for "description logic". This is a form of logic that class construction as the primary modeling mechanism. A class is essentially the same as the notion of set in mathematics. A class is *constructed* by specifying its members using other classes. For example, the definition of autoradiography is "A technique that uses X-ray film to locate radioactively labeled molecules or fragments of molecules." From a DL perspective, autoradiography is a class consisting of those members of the technique class that use X-ray film to locate radioactively labeled molecules or fragments of molecules. Queries to an OWL-DL ontology would mostly be concerned with whether or not a specific entity belongs to a specified class.

Expressing BNs using richer ontology languages, such as RDF or OWL, would be beneficial for a number of reasons. One can take advantage of language constructs that exist in RDF and OWL that cannot be expressed in XML alone. RDF and OWL have inferencing capabilities that XML does not have. A rules language is being developed for OWL. If BNs were expressed using OWL, then it should be possible to specify both logical rules and probabilistic rules in the same document.

5. Combining the Semantic Web with Bayesian Networks

We now give a concrete proposal for how the Semantic Web can be augmented to include BNs and stochastic inference. The architecture for the Semantic Web consists of a series of layers as shown in Figure 2. This figure was taken from a presentation by Tim Berners-Lee[1]. The layers that are relevant to the BW are the following:

(1) The Resource Description Framework (RDF) layer introduces semantics to XML. It makes it possible to link one resource to another resource such that the link and resources may be in different Web pages. RDF is a minimalist semantic layer with only the most basic constructs.
(2) The Web Ontology (OWL) layer expands on the RDF layer by adding more constructs and richer formal semantics.
(3) The Logic layer adds inference. At this layer one can have both resources and links that have been inferred. However, the inference is limited by the formal semantics specified by RDF and OWL.
(4) The Proof layer adds rules. Rules can take many forms such as logical rules as in the Logic Layer, search rules for finding documents that match a query, and domain-specific heuristic rules.

The proposed BW consists of a collection of ontologies that formalize the notion of a BN together with stochastic inference rules. The BW resides primarily on two of the SW layers: the Web Ontology layer and the

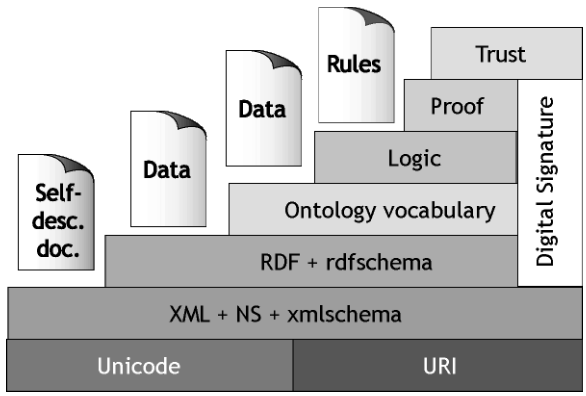

Figure 2. The Semantic Web architecture.

Proof layer. The BW ontologies are expressed in OWL on the Web Ontology layer, and the algorithms for the stochastic operations are located on the Proof layer. By splitting the BW into two layers, one ensures that BW information can be processed using generic SW tools which have no understanding of probability or statistics. The result of processing at the OWL layer is to obtain authenticated and syntactically consistent BNs. The probabilistic and statistical semantics is specified on the Proof layer which requires engines that understand probability and statistics.

6. The Bayesian Web Ontology

The ontology for BNs is built from three sub-ontologies, each of which imports the previous ones:

(1) The ontology of elementary probability distributions.
(2) The ontology of networks of conditional probability distributions.
(3) The ontology of phenomena which can be modeled using BNs.

In this section we construct these ontologies

The top level concept of the BW is the BN which is used to model network of more elementary phenomena. See Figure 3. A BN consists of a collection of *nodes*, each of which represents one elementary phenomenon. Think of a node as a random variable whose probability distribution has not yet been specified. A node has a range of values. For example, the height of a person is a positive real number. A Node can *depend on* other Nodes. A dependency is called a *dependency arc*. It is convenient to order

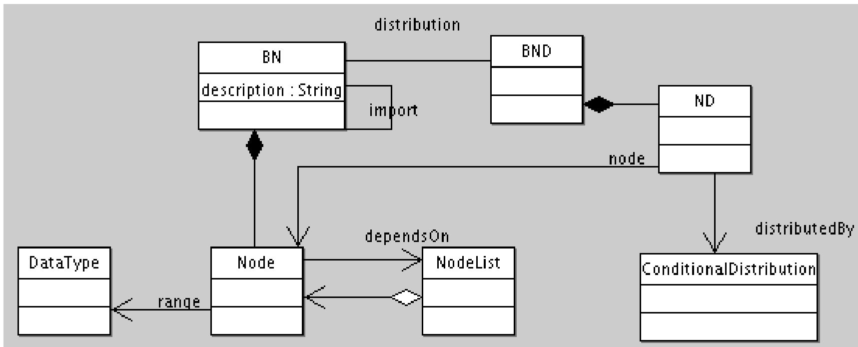

Figure 3. Ontology for Bayesian Networks

the dependencies of a single node, so in Figure 3, a Node an depend on a NodeList, which consists of a sequence of Nodes. The order of the dependencies is used when the conditional probabilities are specified. A BN can *import* another BN. The nodes and dependencies of an imported BN become part of the importing BN.

The most complex part of a BN is its joint probability distribution which is specified using a collection of conditional and unconditional probability distributions. Since a BN can have more than one probability distribution, the notion of a BN distribution (BND) is separated from that of the BN. There is a one-to-many relationship between the concepts of BN and BND. A BND consists of a collection distributions, one for each node in the BN. A node distribution (ND) relates one node to its conditional (probability) distribution.

The notion of a conditional (probability) distribution is the main concept in the conditional probability ontology, as shown in Figure 4. A conditional distribution has three special cases. It can be a *conditional probability distribution table* (CPT), a *general stochastic function* (SF) or an (unconditional) probability distribution. The first of these is used by phenomena with a small number of possible values (called *states* in this case). Most current BN tools support only this kind of conditional probability specification.

A CPT is defined recursively, with one level for each dependency. There is one conditional probability entry (CPE) for each value of the first parent node. Each CPE specifies a weight and a CPT for the remaining parent nodes. Weights are nonnegative real numbers. They are normalized to define a probability distribution. At the last level one uses an unconditional

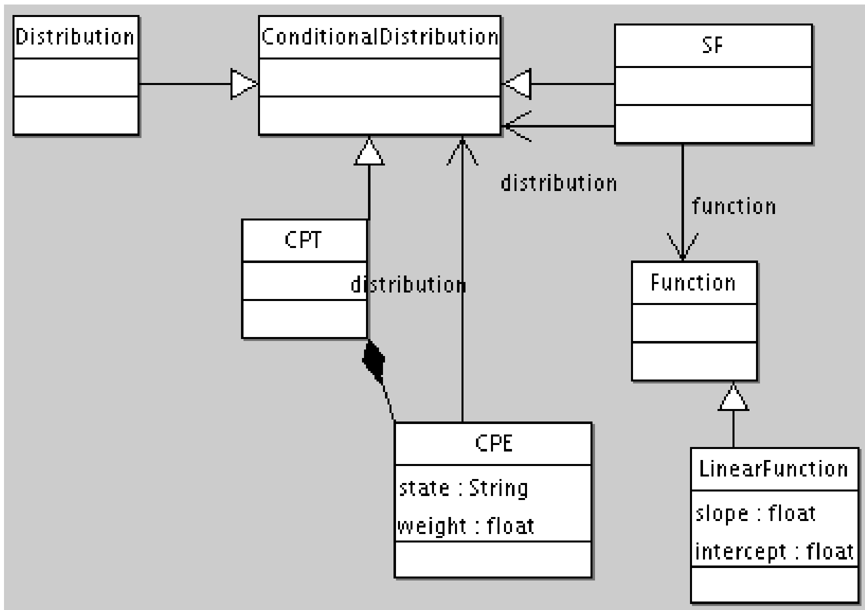

Figure 4. Ontology for Conditional Probability Distributions

probability distribution.

A SF is also defined recursively, but instead of using an explicit collection of CPEs, it uses one or more functions that specify the parameter(s) of the remaining distributions. The most common function is linear function, and it is the only one currently included, but others can be added. This is necessary for dependencies on continuous phenomena.

Probability distributions are classified in the Probability Distribution ontology shown in Figure 5. This ontology is a hierarchy of the most commonly used probability distributions. The main classification is between discrete and continuous distributions. Discrete distributions may either be defined by a formula (as in the Poisson and Binomial distributions) or explicitly for each value (state). Every continuous distribution can be altered by changing its *scale* or by *translating* it (or both). The most commonly used continuous distributions are the uniform and Gaussian (normal) distributions. The uniform distribution is on the unit interval and the Gaussian has mean 0 and variance 1. Other uniform and Gaussian distributions can be obtained by scaling and translating the standard ones. Other commonly used distributions are the exponential and chi-square distributions as well as the Student's t (due to Gosset) and Fisher's F.

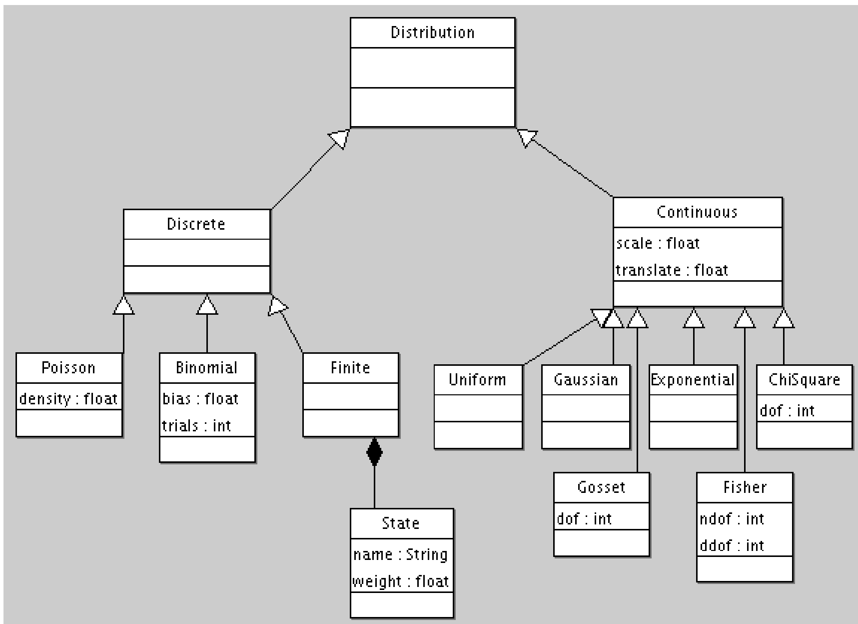

Figure 5. Ontology for Probability Distributions

7. Conclusion

This paper has presented an extension of the Semantic Web that integrates probabilistic inference with logical inference. In the process it opens possibilities for automating processes such as reconciliation, consistency checking and information fusion of scientific results from diverse sources. However, many challenges remain before the BW can be fully realized. As a first step, existing tools for BN analysis must be adapted to use the proposed BW ontology. A more fundamental problem is to specify the semantics of the BW. While there is a formal semantics for the SW and BNs separately, there is no formal semantics that combines the two. At the least, there should be minimum logical requirements for BN information from two sources to be fusable. If it has been determined that this information is fusable, then there should be a formal mathematical definition of the fused result.

8. Acknowledgments

The author would like to thank the Division of Preventive Medicine of The Harvard Medical School for their support and encouragement. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of Harvard University.

References

1. T. Berners-Lee. Semantic Web - XML2000, 2000. `www.w3.org/2000/Talks/-1206-xml2k-tbl`.
2. T. Berners-Lee, J. Hendler, and O. Lassila. The Semantic Web. *Scientific American*, May 2001.
3. N. Friedman. Inferring cellular networks using probabilistic graphical models. *Science*, 303:799, February 6 2004.
4. Glymour and Cooper, editors. *Computation, Causation and Discovery*. MIT Press, 1999.
5. T. Jaakkola and M. Jordan. Variational probabilistic inference and the QMR-DT network. *Journal of Artificial Intelligence Research*, 10:291–322, 1999.
6. D. Koller and A. Pfeffer. Object-oriented bayesian networks. In *Proc. Thirteenth Ann. Conf. on Uncertainty in Artificial Intelligence*, pages 302–313, Providence, RI, August 1997.
7. O. Lassila and R. Swick. Resource description framework (RDF) model and syntax specification, Feburary 1999. `www.w3.org/TR/REC-rdf-syntax`.
8. D. McGuinness. Ontologies and online commerce. *IEEE Intelligent Systems*, 16(1):8–14, 2001.
9. J. Pearl. Graphical models for probabilistic and causal reasoning. In D. Gabbay and P. Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems, Volume 1: Quantified Representation of Uncertainty and Imprecision*, pages 367–389. Kluwer Academic Publishers, Dordrecht, 1998.
10. J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
11. B. Shipley. *Cause and Correlation in Biology*. Cambridge University Press, 2000.
12. M. Smith, D. McGuinness, R. Volz, and C. Welty. OWL web site, November 2002. `www.w3.org/TR/owl-guide/`.
13. Spirtes, Glymour, and Scheines. *Causation, Prediction and Search*. MIT Press, 2001.
14. Microsoft Decision Theory and Adaptive Systems Group. XML belief network file format, April 1999. `research`.microsoft.com/dtas/bnformat/-xbn_dtd.html.