

A method based on genetic algorithms and fuzzy logic to induce bayesian networks

Manuel Martínez Morales, Ramiro Garza Domínguez
Nicandro Cruz Ramírez, Alejandro Guerra Hernández, José Luis Jiménez Andrade
Universidad Veracruzana
Facultad de Física e Inteligencia Artificial
Sebastián Camacho No 5, Xalapa, Ver.,
México 91000

Abstract

A method to induce bayesian networks from data to overcome some limitations of other learning algorithms is proposed. One of the main features of this method is a metric to evaluate bayesian networks combining different quality criteria. A fuzzy system is proposed to enable the combination of different quality metrics. In this fuzzy system a metric of classification is also proposed, a criterium that is not often used to guide the search while learning bayesian networks. Finally, the fuzzy system is integrated to a genetic algorithm, used as a search method to explore the space of possible bayesian networks, resulting in a robust and flexible learning method with performance in the range of the best learning algorithms of bayesian networks developed up to now.

1. Introduction

Bayesian networks are one of the best known formalisms to reason under uncertainty in Artificial Intelligence (AI). A Bayesian network is a probabilistic graphical model that represents a given problem through a graph, in which the nodes represent variables, and the arcs relationships of dependency between these variables.

In order to acquire knowledge from data, different machine learning methods have been proposed. There are two classes of algorithms in the context of learning Bayesian networks from data. Constraint based algorithms try to find out the dependencies and independencies between the variables, implicit in data, in order to induce the structure of the network. Search and scoring based algorithms perform a search in a space of possible network structures, looking for the structure that best fits the data, guided by some metric for scoring. The most used search methods are: greedy hill-

climbing, best-first search, simulated annealing, and genetic algorithms. Metrics for scoring include: Bayesian measure, minimum description length principle, minimum message length, and Akaike information criteria.

We propose a search and scoring learning method, where the key idea is to take advantage of the information provided together by different metrics, instead of exclusively applying only one well known metric. Observe that each one of these metrics is based on different properties, then the point is to evaluate the Bayesian networks using these different properties simultaneously. In order to achieve that, a fuzzy system is proposed to compute the scoring method. The fuzzy system offers a flexible method for combining these metrics, enabling quality evaluation of the network from different perspectives. For the search method, a binary genetic algorithm is used. Genetic algorithms are well suited to explore complex search spaces, and they have already been used to learn Bayesian networks [8]. The genetic algorithm used has been adapted in terms of the metric we propose.

The paper is organized as follows. Section 2 describes how different metrics are integrated into a fuzzy system to evaluate Bayesian networks. Section 3 presents the characteristics of the genetic algorithm used as the search method. Section 4 reports the results obtained with this methodology, using different data bases. Finally, Section 5 includes conclusions and future work.

2. Fuzzy evaluation

The proposed metric combines four different metrics in a fuzzy system: Bayesian measure, the minimum description length principle, Akaike information criteria, and estimated classification accuracy. The first three are usually applied in existing algorithms. The estimated accuracy, defined as the percentage of examples correctly classified by

the network, usually is not applied to guide the search because of its computational cost.

The minimum description length (MDL) principle and the Akaike information criteria are very similar. Both of them use the concept of conditional entropy related to the structure of a Bayesian Network [2]. The entropy is a non negative measure of uncertainty, maximal when total uncertainty is present, and zero under complete knowledge. When information is increased, the entropy decreases. This means that adding arcs to the network reduces entropy, because the probability distribution is better described by adding this arc. However, using exclusively entropy to guide the search, introduces a bias in favor of more complex networks, i.e., densely connected networks. So, a property penalizing complexity is needed. The MDL and the Akaike information criteria have two terms, one measuring entropy, and another controlling complexity. They differ on the second term. Networks with small values for these metrics are preferred.

The Bayesian measure[3] tries to maximize the probability of the structure of the network Bs , given the data base D , $P(Bs|D)$, over all the possible network structures. P is calculated for a set of network structures, and the network with higher conditional probability is selected.

Estimated accuracy is based on the performance of the network classifying data. The main idea is to approach the Bayesian network as a classification model, selecting one variable as the target class, and trying to learn a mapping from the other variables to the target variable. The first step in order to achieve learning is to train the Bayesian network with a subset of the data, i.e., calculate the conditional probability tables of the whole network. The estimated accuracy, i.e., the proportion of correctly classified instances, is calculated over the remaining data.

The proposed metric is defined using these four metrics as base variables over which, linguistic variables are defined in a fuzzy system. Hence, the system consists of five variables: length, associated to the MDL principle; information, for the Akaike information criteria; probability, for the Bayes measure; classification, for the estimate accuracy; and quality, as the output variable expressing the evaluation of the network. Input variables use three linguistic terms (low, medium, high), while the output variable uses five (very bad, bad, medium, good, excellent). Each linguistic term is defined by a fuzzy set, defined in turn by a membership function. First, we define the domain of the membership functions for the MDL principle and the Akaike information criteria. Observe that, when starting with an empty network, the values of these metrics decrease everytime an arc is added, but eventually the values will increase if we continue adding arcs. A theoretical global minimum exists between 0 and an upper bound defined as the higher value of the metric for the empty and the full connected networks.

To define the membership functions for the Bayes mea-

sure, a different criterium is used. In principle, the values of this metric must be in the range $0 \dots 1$, but in practice, using the original definition of this measure produces numeric overflow. So, it is necessary to use the logarithm of the probabilities. The values generated in this way are negative reals. Higher probabilities converge to zero. The upper bound for this membership function is zero, and the lower bound is the smaller value of the modified measure for the empty and the full connected networks. For the estimated accuracy, it is clear that the range of the measure is in $0 \dots 1$.

For the proposed metric, the membership functions are estimated by fuzzy inference over the other four metrics. All the fuzzy sets, for all the linguistic variables are distributed uniformly on the respective ranges. The membership functions Z and S are used for the extremes, and the lambda function is used for middle fuzzy sets. Figure 1 shows the fuzzy sets for the quality output variable in the combined metric.

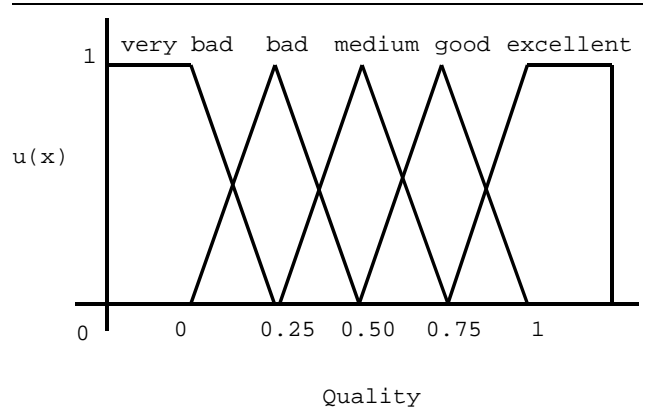


Figure 1. Fuzzy sets for the quality variable.

Now it is possible to define the fuzzy rules that model the decision process about the quality of a Bayesian network. For simplicity, we only considered rules with one variable in the antecedent. Two linguistic variables are to be maximized: Bayes measure and estimated accuracy. Two linguistic variables are to be minimized: MDL and Akaike information criteria. The fuzzy rules are shown in table 1. The truncating method is used for the output set. For aggregation, addition is used.

3. Genetic search

The approach adopted to search the space of hypotheses is to use a genetic algorithm. These algorithms are inspired in natural selection and genetics, operating on a population of individuals called chromosomes. Individuals are vectors

r1:	length = low → quality = excellent
r2:	length = medium → quality = good
r3:	length = high → quality = medium
r4:	probability = low → quality = medium
r5:	probability = medium → quality = good
r6:	probability = high → quality = excellent
r7:	information = low → quality = good
r8:	information = medium → quality = medium
r9:	information = high → quality = bad
r10:	clasification = low → quality = very bad
r11:	clasification = medium → quality = bad
r12:	clasification = high → quality = medium

Table 1. Fuzzy rules

in the hyperspace being explored. They represent possible solutions to the problem. In our approach, the solution corresponds to the structure of a Bayesian network that maximizes the proposed metric. The structure of the network must be codified as an arrangement of bits. The representation in our system, originally proposed by Larrañaga [8], codes the network using a connectivity matrix, where each row corresponds to a node in the network, and each column indicates with a one, if an arc from the node labeling the column to the node labeling the row exists. The absence of an arc is represented by zero. The connectivity matrix is transformed to a one dimensional array, concatenating the rows. The fitness function is the fuzzy metric proposed above. Genetic operators, e.g., selection, crossover, and mutation, are applied as suggested by Haupt [5]. Generations are computed by rank selection (sorting individuals by fitness) with probability 0.5. Mutation is applied with probability 0.01. Figure 2 shows the genetic algorithm (search module) interacting with the fuzzy system to induce Bayesian networks.

4. Experimental results

In the context of learning Bayesian networks, there are two approaches to evaluate the quality of the induced network. The first consists in comparing the induced network against a golden network. A golden network is handcrafted by experts in the domain of application, so its structure and parameters are known. Using this golden network, a data base is generated using the Monte Carlo method. The data base obtained in this way is used as input for our learning method. The resulting network is compared against the golden network to establish its “goodness of fit”. The second approach consists in using estimated accuracy of clasification for the induced network.

The first experiment is performed on the Asia golden network [7], a fictitious problem of medical diagnosis, about whether a patient has tuberculosis, lung cancer or bronchi-

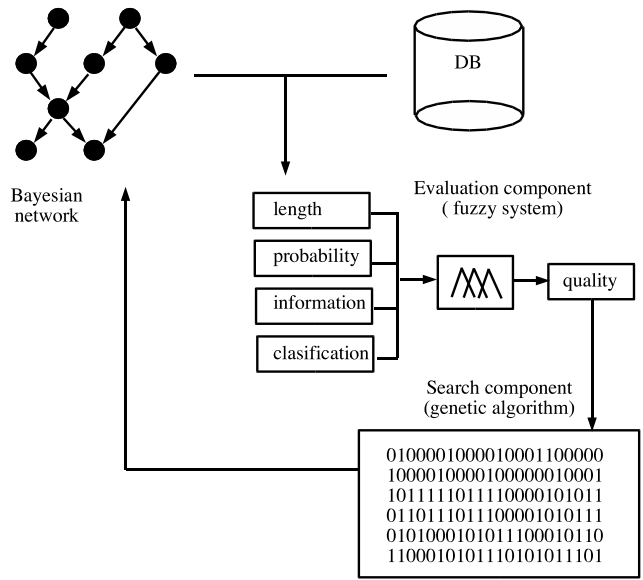


Figure 2. Our genetic-fuzzy system to induce Bayesian networks.

tis, related to their X-ray, dyspnea, visit-to-Asia and smoking status. This golden network has 8 binary variables, and 8 arcs (see Fig. 3).

The second experiment is performed on the Alarm [1] golden network. This network is part of a system for monitoring of intensive care patients. It has 37 variables, and 46 arcs. Among the variables, 8 of them represent diagnostic problems, 16 represent conclusions, and 13 represent intermediate variables. Each variable can have 2 up to 4 values.

The third experiment is performed on the College data base, a.k.a., Sewell & Shah [10], who investigated factors that influence the intention of high school students to attend college. The following variables for 10,318 Wisconsin high school seniors, were measured: Sex (male, female), Socioeconomic Status (low, lower middle, upper middle, high), and Intelligence Quotient (low, lower middle, upper middle, high). Since no golden network has been proposed for this problem, the reference is the network generated by the Heckerman algorithm [6].

Figure 2 shows comparative results for these three experiments, comparing our method (genetic-fuzzy) against four other algorithms: K2 [3], Tetrad [11], Bayes9 [4], BayesN [9]. The first two of them are well known algorithms reported in literature. Bayes9 and BayesN are alternative approaches developed by our team. The column labeled *total*, displays the number of arcs induced by these algorithms. The column *correct* is for the number of arcs placed correctly after the golden network. The column *extra* is for the

Asia	Total	Correct	Extra	Absent
Genetic-Fuzzy	9	8	1	0
K2	8	7	1	1
Tetrad	4	4	0	4
Bayes9	4	4	0	4
BayesN	8	5	3	3

Alarm	Total	Correct	Extra	Absent
Genetic-Fuzzy	46	46	0	0
K2	46	45	1	1
Tetrad	49	46	3	0
Bayes9	44	43	1	3
BayesN	45	32	13	14

College	Total	Correct	Extra	Absent
Genetic-Fuzzy	7	7	0	0
K2	7	7	0	0
Tetrad	7	7	0	0
Bayes9	7	7	0	0
BayesN	7	7	0	0

Table 2. Comparative experimental results against 4 algorithms, using Asia, Alarm, and College databases

number of arcs induced by the algorithm which did not appear in the golden network. Finally, *absent* stands for the number of arcs present in the golden network which did not appear in the induced network.

As results show, our method has the best performance among the tested algorithms. This suggests that the information of different metrics combined in the fuzzy component of our system, exploits better the dependency relationships implicit on data. Figure 3 shows graphically, as an example, the Asia golden network and the networks for this problem generated by the tested algorithms.

5. Conclusions and future work

Learning Bayesian networks from data is an open problem. The method proposed in this paper, instead of applying an isolated well known metric, takes advantage of the information that different metrics could provide together, combined in a fuzzy system. A genetic algorithm uses this information to search the space of possible structures for a network.

The experimental results show that the performance of the method proposed, compares well the best algorithms reported in literature with respect to learning golden networks.

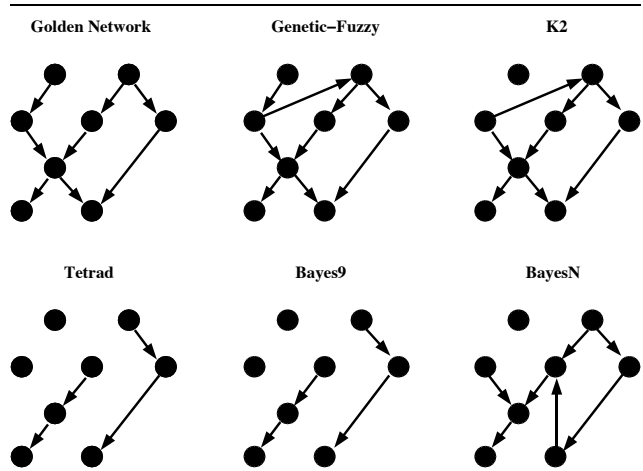


Figure 3. Asia Bayesian network.

Some of the limitations of the method include: i) Evaluating the estimated accuracy of classification is computationally expensive; ii) The method is sensitive to selection of initial population, as all genetic algorithms; iii) It might be that in some problems combination of just two metrics could be appropriate; and iv) Sensitivity to different membership functions remains to be assessed.

In respect to the fuzzy component of the method, future work includes: i) Use different fuzzy sets; ii) Use increasingly complex fuzzy rules; iii) Include linguistic quantifiers; and iv) Consider more advanced fuzzy operators. In respect to the genetic component, different selection methods can be tested. Finally, the idea of using a fuzzy system to evaluate hypothesis combining different metrics, while searching with a genetic algorithm, can be applied beyond Bayesian networks, e.g., decision trees.

References

- [1] Beinlich, I., Suermondt, G., Chavez, R., and Cooper G.: The alarm monitoring system. In: Second European Conference on AI and Medicine, Berlin. Springer-Verlag, Germany (1989)
- [2] Boukaert, R.R.: Probabilistic network construction using the minimum description length principle. Technical Report, Ulrecht University (1994)
- [3] Cooper, G.F. and Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9, 309–347 (1992)
- [4] Cruz Ramírez, N.: Building Bayesian networks from data: a constraint based approach. PhD Thesis. The University of Sheffield (2001)
- [5] Haupt, R.L. and Haupt S.E.: *Practical Genetic Algorithms*. John Wiley and Sons Inc., New York, USA (1998)
- [6] Heckerman D.: A tutorial learning with bayesian networks. *Learning in graphical models*, M. Jordan, (ed.) MIT Press, Cambridge, MA, USA (1999)

- [7] Lauritzen, Steffen L. and David J. Spiegelhalter.: Local computations with probabilities on graphical structures and their application to expert systems. In: J. Royal Statistics Society B, 50(2), 157-194 (1988)
- [8] Larrañaga, P.: Structure Learning of Bayesian Networks by Genetic Algorithms: A Performance Analysis of Control Parameters. IEEE Journal on Pattern Analysis and Machine Intelligence, 18(9), 912–926, USA (1996)
- [9] Martínez Morales, M., Cruz Ramírez, N., Jiménez Andrade, J.L., and Garza Domínguez, R.: Bayes-N: an algorithm for learning bayesian networks from data using local measures of information gain applied to classification problems. In: MICAI 2004: Advances in Artificial Intelligence, Monroy, R.; Arroyo-Figueroa, G.; Sucar, L.E.; Sossa, H. (Eds.) Lecture Notes in Artificial Intelligence, Vol.2972, Springer Verlag, Germany (2004)
- [10] Sewell, W. and Shah, V. (1968). Social class, parental encouragement, and educational aspirations. American Journal of Sociology, 73:559-572.
- [11] Spirtes, P.R., Shchepines, R. and Clark, G.: TETRAD II: Tools for Discovery. Lawrence Erlbaum Associates, Hillsdale, NJ., USA (1994)