Journal of Computer Science

Original Research Paper

MCMC-Fuzzy: A Fuzzy Metric Applied to Bayesian Network Structure Learning

[1]Ademar Crotti Junior, [2]Beatriz Wilges and [3]Silvia Modesto Nassar

[1]*School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland*
[2]*Department of Information Systems, Federal University of Santa Catarina, Florianópolis, Brazil*
[3]*Department of Informatics and Statistics, Federal University of Santa Catarina, Florianópolis, Brazil*

Article history
Received: 02-04-2018
Revised: 25-05-2018
Accepted: 10-08-2018

Corresponding Author:
Ademar Crotti Junior
School of Computer Science and Statistics, Trinity College Dublin, Dublin, Ireland
Email: crottija@tcd.ie

Abstract: Bayesian network structure learning is considered a complex task as the number of possible structures grows exponentially with the number of variables. Two main methods are used for Bayesian network structure learning: Conditional independence, a method in which a structure is created consistently with independence tests performed on data; and the heuristic search method that explores the structure space. Hybrid algorithms combine both of the aforementioned methods. In this study, we propose the combination of common metrics, used to evaluate Bayesian structures, into a fuzzy system. The idea being that different metrics evaluate different properties of the structure. The proposed fuzzy system is then used as a metric to evaluate Bayesian networks structures in a heuristic search algorithm based on Monte Carlo Markov Chains. The algorithm was evaluated within the context of synthetic databases through comparison with other algorithms and processing time. Results have shown that, despite an increase in processing time, the proposed method improved the structure learning process.

Keywords: Fuzzy Systems, Bayesian Network Learning, Markov Chain Monte Carlo (MCMC)

Introduction

Bayesian networks are probability models that represent knowledge under random uncertainty. They are used in several areas, such as in behavior predictions, natural language processing, robotics, among others (Friedman and Koller, 2003).

Bayesian networks are composed of two main components: Parameters and structure. Parameters define the conditional probabilities between the variables or nodes. Structure defines the network topology, where the connections and the direction of such must be determined (Castillo *et al*., 2012).

Often domain experts are needed to define Bayesian network structures. However, this can be costly, complex and time-consuming due to the amount of variables, incompleteness of data and the difficulty in maintaining the structure, making the process impracticable in several cases (Scutari and Denis, 2014). Therefore, significant research effort has been invested into learning Bayesian networks structure from data. Both components can be learned from data. The learning of parameter is considered a simple task when the structure of the network is well defined. On the other hand, structure learningis an NP-Hard problem (Chickering, 1996).

The literature review shows two main methods to learn Bayesian network structures. The first method uses conditional independence tests on the data to find a structure consistent with the observed independence. The problem with this approach is the exponential number of dependence tests (Margaritis and Thrun, 1999). The second method defines a function to evaluate how well a structure represents the data, finding the simplest structure that increases the value of this function (Ko and Kim, 2014). Algorithms in this method explore the search space, assessing the structures score metric functions. A problem with this approach is that the search space contains all the possible structures (Yan and Cercone, 2010). Some algorithms, named hybrid, combine both approaches (Margaritis and Thrun, 1999). The most common approach in hybrid methods is to use independence tests to restrict the search space and then apply a heuristic search method (Tsamardinos *et al*., 2006).

Science Publications

The research question posed for this work is: *To what extent the evaluation of different properties of Bayesian networks improve the structure learning process in a heuristic search algorithm?* To tackle this problem, we propose a fuzzy system, which combines common evaluation metrics. This approach draws inspiration from (Morales *et al*., 2004). Fuzzy systems are flexible tools capable of approximating different functions into one through a set of rules (Brooks *et al*., 2011). Moreover, fuzzy systems have been widely used in many fields, including in the Bayesian networks domain, for learning structure (Morales *et al*., 2004) and inference (Yang, 1997).

The main contribution of this work is the development of a fuzzy system that combines different common metrics to evaluate Bayesian network structures. The metric is then applied to a Markov Chain Monte Carlo search algorithm in the process of learning Bayesian network structures. Even though this work applies the metric to an MCMC algorithm, the metric is generic enough to be applied to other search algorithms. The proposed approach is evaluated by comparing its results to the results of other state-of-the-art algorithms in the context of two synthetic databases.

The article is thus presented as follows: Section II presents the related work. Section III presents the background knowledge on Bayesian network structure learning. Section IV presents the proposed fuzzy metric that is applied to a Markov Chain Monte Carlo algorithm. Section V presents the evaluation and Section VI concludes the article.

Related Work

Bayesian networks are widely used and thus automatic learning these networks from data is a very active research field (Chickering, 2003; Tsamardinos *et al*., 2006; Friedman and Koller, 2003; Vafaee; 2014; Guo and Li, 2009). This literature review considered articles published in digital libraries. The databases used in this review were IEEE Xplore, ACM digital Library, Springer Link and ScienceDirect. The selected articles are presented below. These were selected based on their importance as well as current trends in the field. The review considered papers published between 1995 and 2017.The research presented in this paper focuses on heuristic search algorithms. Therefore, the articles presented in this Section fall under the heuristic search approach.

Table 1 presents a summary of the approaches analyzes in this paper. Many of these approaches apply genetic and greedy algorithms to learning Bayesian network structures. Greedy algorithms find an optimal local, hoping that a local optimum also represents the global optimum of the problem. Greedy algorithms usually have good performance, however, they are highly dependent of its initial stage. The K2 (Heckerman *et al*., 1995) algorithm is one of the most known greedy algorithms in Bayesian structure learning. The main drawback of the K2 algorithm is that the order of the nodes is required as a parameter. Ko and Kim (2014) proposed an algorithm to define this parameter. Chickering (2003) proposed the Greedy Equivalence Search (GES) algorithm, which has gained many extensions. Nielsen *et al*. (2002) added randomness to this algorithm that attempts to escape from local optima. The extension of the GES proposed by Alonso-Barba *et al*. (2011) aimed at improving its performance. Scanagatta *et al*. (2015) proposed a greedy algorithm to work on large networks. Genetic algorithms are inspired by natural evolution, where the fittest individuals are selected in each iteration, to produce the next sample population. Examples of such algorithms include a genetic algorithm which also uses a fuzzy system (Morales *et al*., 2004). Another genetic algorithm applies the K2 scoring function to evaluate the produced structures (Faulkner, 2007). There is also a genetic algorithm focused on learning large structures (Vafaee, 2014).

Hybrid algorithms combine heuristic search and conditional independence tests to the data. For example, in (Zhang *et al*., 2013), in order to limit the search space, the authors proposed the use of conditional independence tests to construct an undirected graph. The resulting graph is applied as input to a greedy algorithm that determines the direction of the relations. See Table 1 for other algorithms that apply different approaches to limit the search space of Bayesian structures.

Table 1: Summary of Bayesian network structure learning methods

Approach	Authors
Genetic algorithms	Morales *et al*. (2004) Faulkner (2007) Vafaee (2014)
Greedy algorithms	Heckerman *et al*. (1995) Chickering (2003) Nielsen *et al*. (2002) Alonso-Barba *et al*. (2011) Ko and Kim (2014) Scanagatta *et al*. (2015)
Hybrid algorithms	Tsamardinos *et al*. (2006) Gámez *et al*. (2011) Yuan and Malone (2013) Zhang *et al*. (2013) Xu and Srihari (2014)
Markov Blanket algorithms	Margaritis and Thrun (1999) Pellet and Elisseeff (2008) Aliferis *et al*. (2010) Zhu and Yang (2014) Sechidis and Brown (2015)
MCMC algorithms	Friedman and Koller (2003) Grzegorczyk and Husmeier (2008) Guo and Li (2009) Niinimaki *et al*. (2012) Masegosa and Moral (2013) Su *et al*. (2014) Su and Borsuk (2016)

Markov blanket algorithms use techniques to identify the Markov blanket of each possible node. A Markov blanket of a given variable X is defined as the subset of variables that if observed, make the variable X conditionally independent of all other variables (Margaritis and Thrun, 1999). These algorithms propose different approaches to define the Markov Blanket, which is then used to define the Bayesian network structure (Margaritis and Thrun, 1999; Aliferis *et al*., 2010; Sechidis and Brown, 2015). Existing algorithms include the work of Pellet and Elisseeff (2008), which proposed the use of feature selection algorithms to define Markov blankets. Zhu and Yang (2014) decomposes the BN structure using Markov blankets and then defines the orientation of the arcs using the K2 scoring function.

Markov Chain Monte Carlo (MCMC) algorithms are methods that explore the search space by sampling from a probability distribution. MCMC methods apply correlated random sampling to move around the chain (Friedman and Koller, 2003). Examples of MCMC algorithms include its use to find the order of the nodes as a first step, then its application to search for the structure of the network using the order found previously (Friedman and Koller, 2003). Grzegorczyk and Husmeier (2008) proposed the remove arc modification operation into the MCMC method. Guo and Li (2009) combined the Expectation–Maximization algorithm to the MCMC algorithm. Niinimaki *et al.* (2012) proposed another algorithm over the order of the nodes, similar to the one proposed by Friedman and Koller (2003). Masegosa and Moral (2013) combines stochastic search and MCMC sampling. Su *et al.* (2014) proposed the incorporation of external knowledge into the MCMC algorithm. Su and Borsuk (2016) used Markov blanket resampling as a step of the MCMC algorithm.

As mentioned before, Table 1 summarizes the main approaches applied in Bayesian network structure learning. The table also shows how the MCMC method is being used in recent and relevant studies for the task of Bayesian network structure learning.

As mentioned before, the proposed fuzzy metric draws inspiration from a genetic algorithm proposed by Morales *et al.* (2004), which also uses a fuzzy metric to evaluate the structures. In this work, we propose a similar fuzzy metric applied to an MCMC heuristic search algorithm in the process of learning Bayesian network structures from data.

Bayesian Network Structure Learning

Bayesian networks can be modeled by domain experts or learned from data. As mentioned before, the model of Bayesian networks by experts can be costly, complex and time-consuming due to the number of variables, incompleteness of the data, amongst others. Structure learning generates a structure based on evidence found in the data. There are two main approaches in Bayesian network structure learning.

The method based on conditional independence treats each and every variable independently of its non-descendants, conditioned to their parents. This means that each and every Bayesian network represents a function of density and unique probability that can be factored (Carvalho and Chiann, 2013).

The second most common method found in the literature review applies heuristic search methods to generate a structure that best represents the data based on a score metric criteria (Daly *et al*., 2011). K2 is an example of an algorithm that applies heuristic search in the process of learning Bayesian structures from data. Some of the main metrics used in algorithms that apply heuristic search methods are: AIC, MDL and BDe.

The Akaike Information Criterion (AIC) is based on two terms: One term controlling entropy, based on conditional entropy and another one controlling complexity of the structure. In Information Theory, entropy is a non-negative value that measures uncertainty, tending to zero when knowledge is high. The AIC metrics is given by (Akaike, 1974):

$$AIC = H\left(G, D\right) + K$$

Considering that:

$$H\left(G, D\right) = -N \sum_{i=1}^{n} \sum_{j=1}^{q_i} \sum_{k=1}^{r_i} \frac{N_{ijk}}{N} \log \frac{N_{ijk}}{N_{ij}}$$

The Minimum Description Length (MDL) metrics uses the same terms of the AIC metric with a small difference in the second term. This metrics is known for finding Bayesian network structures that are simpler than through AIC. The MDL metrics is given by (Bouckaert, 1993):

$$MDL\left(G, D\right) = H\left(G, D\right) + \frac{K}{2} \log N$$

The Bayesian Dirichlet equivalence (BDe) maximizes structure likelihood according to the data - that is, the metric uses the conditional likelihood of each variable in the network. The BDe metrics is given by (Heckerman *et al*., 1995):

$$BDe\left(G, D\right) = P(G) \prod_{i=0}^{n} \prod_{j=1}^{q_i} \frac{\Gamma(r_i)}{\Gamma\left(r_i + N_{ij}\right)} \prod_{k=1}^{r_i} \frac{\Gamma\left(r_i + N_{ijk}\right)}{\Gamma\left(r_i\right)}$$

The AIC metric analyzes network information. The MDL metric analyzes complexity, while the BDe is based on probability.

Fuzzy Metric

In this section, a fuzzy metric that combines different common metrics for structural learning of Bayesian network is proposed. The proposed score fuzzy metric is applied to a heuristic search Markov Chain Monte Carlo method, which is being called MCMC-Fuzzy. The membership functions and fuzzy rules were based on the properties of the metrics being used by the fuzzy system and by attempting different configurations. These were also based on the work of Morales *et al*. (2004).

Monte Carlo via Markov Chains (MCMC) is a method that uses approximation by sampling. The goal is to generate a Markov chain limited by a desired distribution. The method obtains random probability distribution samples that are considered difficult to sample directly. Starting from any point, as the number of samples increases, it is said that the chain is closer to its balance distribution (Brooks *et al*., 2011).

The workflow of the MCMC-Fuzzy algorithm is presented in Fig. 1.

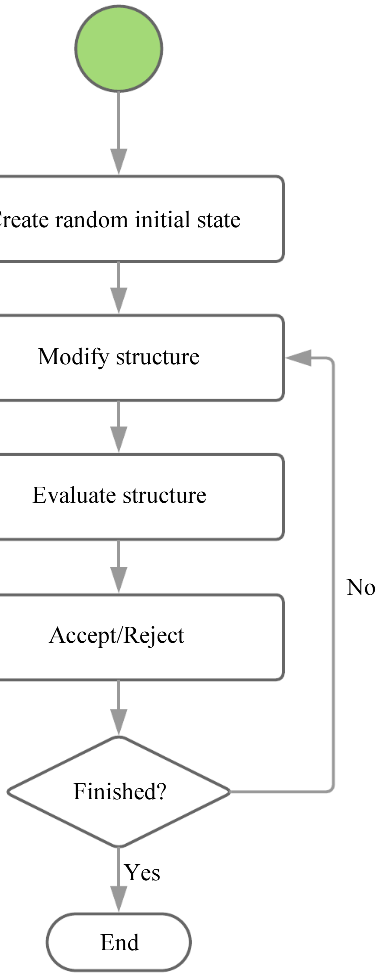


Fig. 1: MCMC algorithm workflow

Given a set of data, the first stage of the algorithm generates a random Bayesian networks structure, which is used as the initial state of the Markov chain. In the next step, two variables are randomly sampled from the set of variables from the dataset. To select a structure modification, a random number is generated from a uniform distribution. The structure modifications can be: Adding, removing, or inverting an arc, between the two sampled variables. Since a Bayesian network structure cannot have cycles, before evaluating the structure, the algorithm validates that.

Usually, Bayesian network structures are evaluated using a single metric, which is the case of the EM-MCMC algorithm. It is known that score metrics evaluate Bayesian structures differently, for example, some metrics favor more complex structures, while others prefer structures with less parameters (Su and Borsuk, 2016). In this work, we present an approach that combines known metrics through a fuzzy system to evaluate Bayesian networks structure. This metric can be used to evaluate the structures being induced by any algorithm that have an evaluation step.

Fuzzy logic provides a way of combining such distinct metrics into one. The metrics used in the proposed fuzzy system are: AIC, MDL and BDe. The fuzzy metric will thus have four variables: 3 inputs and an output variable, named Quality (Table 2).

Each set is defined as a membership function. Membership functions were defined by using the highest value among all metrics, considering a network structure completely connected and the same structure when completely disconnected. All sets were defined uniformly in the [0,1] interval.

Figure 2 presents the membership function for the metrics, in which x represents the value of metrics and $\mu(x)$ its membership.

The membership function for the Quality variable is presented in Fig. 3.

The following step of the fuzzy system was the definition of the fuzzy rules. The MDL and BDe metrics are minimized and AIC is maximized. Thus, the BDe metric influences the Poor, Average and Good fuzzy sets of the Quality variable. MDL and AIC metrics influence the Average, Good and Excellent fuzzy sets of the same output variable. Figure 4 presents the rules used in the proposed fuzzy metric.

Table 2: Fuzzy sets

Variables	Sets
AIC	Low, Moderate, High
MDL	Low, Moderate, High
BDe	Low, Moderate, High
Quality	Poor, Average, Good, Excellent

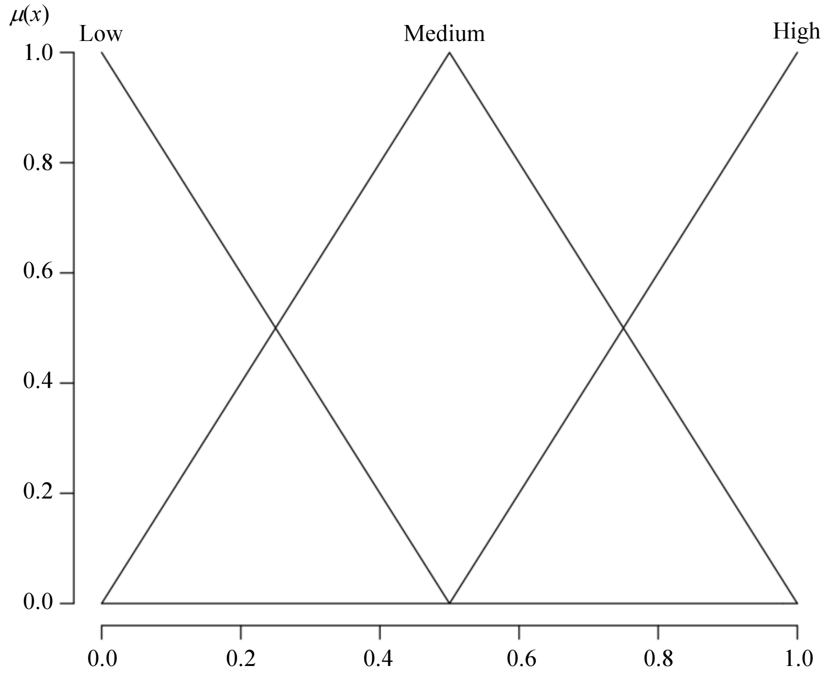


Fig. 2: Membership function for the distinct metrics

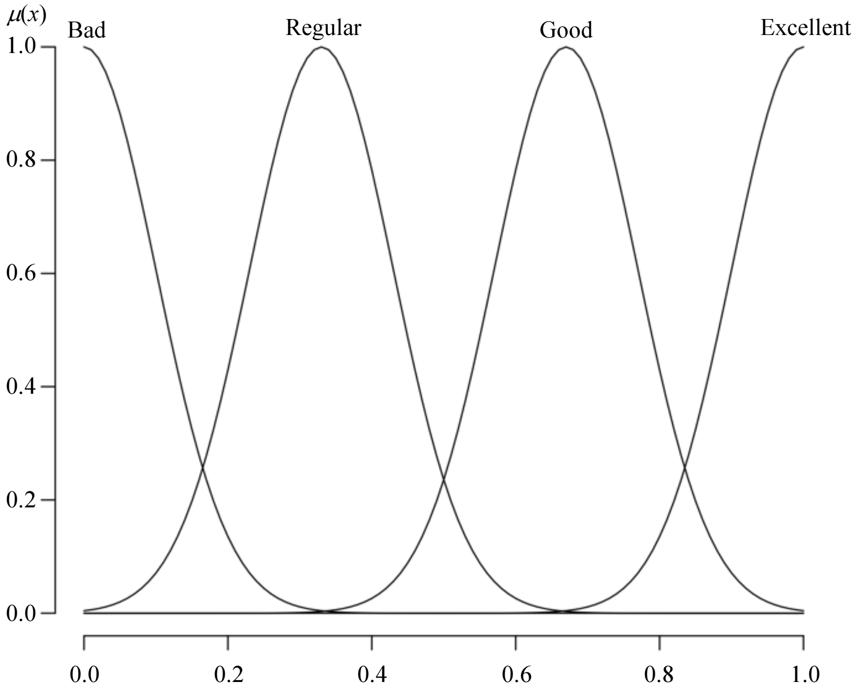


Fig. 3: Membership functions for the output variable Quality

The metrics are aggregated to the Quality variable by truncation. The final part is to defuzzify the Quality output variable. The proposed approach uses the centroid method for defuzzification. The algorithm continues until the amount of iterations is met. The algorithm pseudocode, which is based on the MCMC Metropolis-Hastings algorithm (Chib and Greenberg, 1995), is presented in Fig. 5.

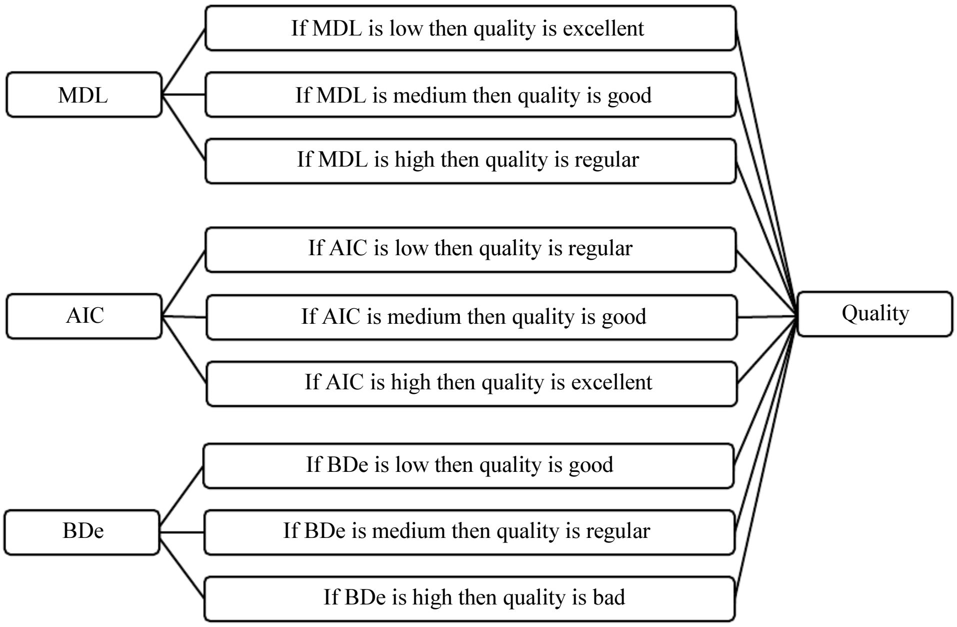
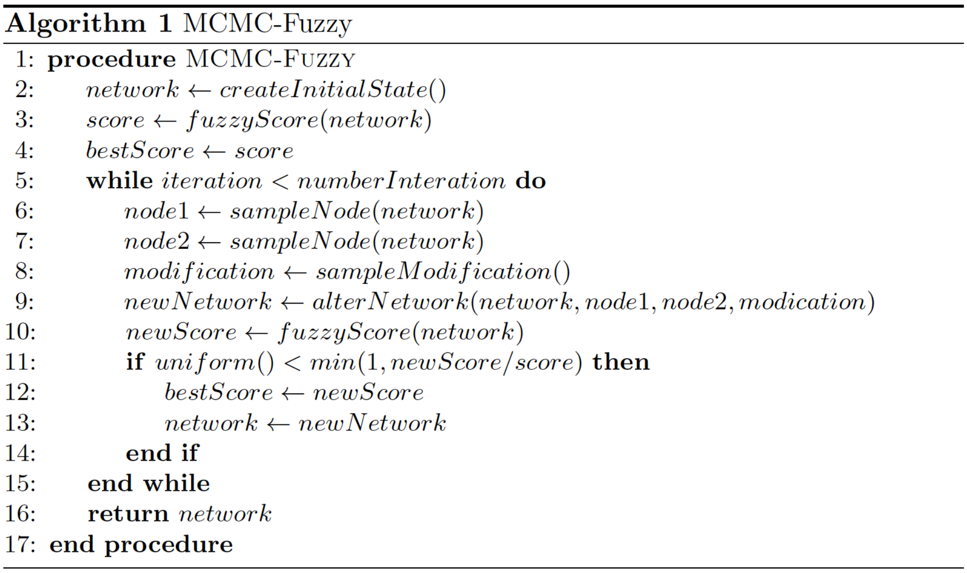

MDL
AIC
BDe
If MDL is low then quality is excellent
If MDL is medium then quality is good
If MDL is high then quality is regular
If AIC is low then quality is regular
If AIC is medium then quality is good
If AIC is high then quality is excellent
If BDe is low then quality is good
If BDe is medium then quality is regular
If BDe is high then quality is bad
Quality

Fig. 4: Rules for the fuzzy metric



Algorithm 1 MCMC-Fuzzy

```
1: procedure MCMC-Fuzzy
2:     $network \leftarrow createInitialState()$
3:     $score \leftarrow fuzzyScore(network)$
4:     $bestScore \leftarrow score$
5:     while $iteration < numberInteration$ do
6:         $node1 \leftarrow sampleNode(network)$
7:         $node2 \leftarrow sampleNode(network)$
8:         $modification \leftarrow sampleModification()$
9:         $newNetwork \leftarrow alterNetwork(network, node1, node2, modication)$
10:        $newScore \leftarrow fuzzyScore(network)$
11:        if $uniform() < min(1, newScore/score)$ then
12:            $bestScore \leftarrow newScore$
13:            $network \leftarrow newNetwork$
14:        end if
15:    end while
16:    return $network$
17: end procedure
```

Fig. 5: MCMC-Fuzzy algorithm pseudocode

The algorithm starts by generating a random structure. Line 3 calculates the initial network score using the proposed fuzzy metric. Line 4 stores this value as the best score. Then, the iterative process starts.

Lines 6 and 7 select 2 nodes randomly. Line 8 specifies one of the possible modifications (add, remove, or invert arcs). The modification is performed in Line 9. Line 10 calculates the score for the new network with the random modification. Line 11 verifies if a random number of a uniform distribution is smaller than the minimum between 1 and the exponential difference between the scores. If it is, the structure is accepted as the new best structure sample.

The test presented in Line 11 represents the probability of chain acceptance according to the equation presented below (Friedman and Koller, 2003):

$$P(x,x') = \min\left(1, \frac{\pi(x')q(x \mid x')}{\pi(x)q(x' \mid x)}\right)$$

Evaluation

The MCMC-Fuzzy was applied to two synthetic databases. These synthetic databases are commonly used in evaluating Bayesian networks learning algorithms, which makes a comparison between methods possible. The databases are also freely available.

The first synthetic database, called Asia, represents patient diagnosis in a hospital emergency room. The Asia dataset has 5000 rows. The Asia gold standard network has 8 nodes, 8 arcs and 18 parameters. The second database is called Alarm. The Alarm dataset has 20000 rows. The Alarm gold standard network has 37 nodes, 46 arcs and 509 parameters. Both datasets are composed of categorial data, which iscommon for Bayesian networks. We refer the reader to Scutari and Denis (2014), for a full description of these databases.

The evaluation conducted compared the structures learned and the processing time from the proposed method to the EM-MCMC, MMHC and K2 algorithms. EM-MCMC is used in this experiment for being an MCMC method, easily comparable to the proposed Fuzzy-MCMC algorithm. MMHC and K2 are used for widely known and used algorithms for learning Bayesian network structures.

Network Structure Comparison

In this section, we present a comparison of the Bayesian network structures learned from the proposed MCMC-Fuzzy algorithm, to the EM-MCMC, MMHC and K2 algorithms. The algorithms used for comparison were selected due to popularity, efficiency and broad utilization.

Figure 6 presents the structure learned from applying the MCMC-Fuzzy algorithm for the Asia database. The outlined arcs represent arcs correctly identified by the algorithm.

Figure 7 presents the Bayesian network learned by the MCMC-Fuzzy algorithm for the Alarm database.

Tables 3 and 4 present the results found by the K2, MMHC, EM-MCMC as well as by the MCMC-Fuzzy algorithm for the Asia and Alarm databases. The EM-MCMC and MCMC-Fuzzy have the number of iterations as their stop condition. Therefore, tests were performed using different values for this parameter. The gold standard presents the correct Bayesian network settings for the database.

The K2 algorithm is widely known for having good results in several databases. This is due to the input parameter defining the order of variables in the Bayesian network. The definition of this parameter is considered a complex task (Friedman and Koller, 2003). Moreover, structure learning is commonly done when there is little or no information about the data. The remaining algorithms learn the network structure without the need of additional parameters, which makes learning process more difficult.

The MMHC algorithm is used in several applications particularly because it can be scaled, with proper run times even when working with a high number of variables. When applied to the Asia and Alarm databases, the algorithm identified 4 correct arcs for Asia and 18 for Alarm. The MMHC results show certain equivalence with EM-MCMC results, identifying nearly the same amounts of correct, extra and missing arcs in both databases.

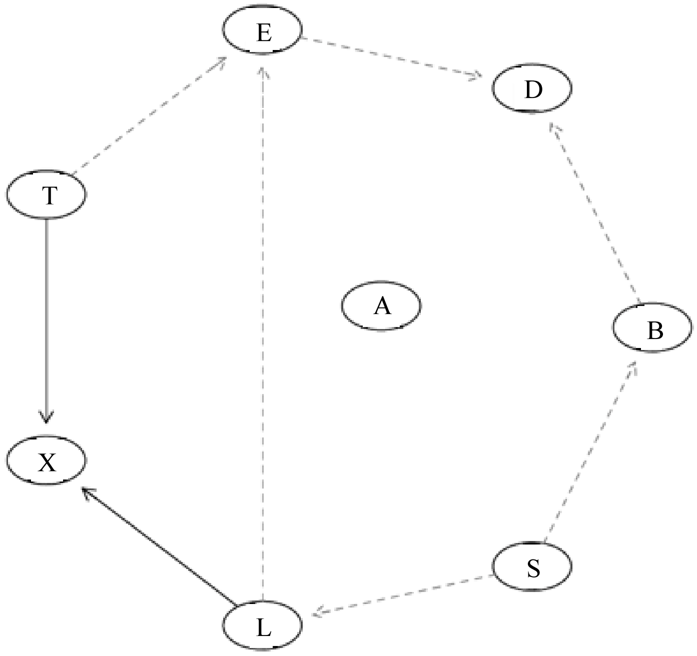


Fig. 6: MCMC-Fuzzy structure for the Asia database

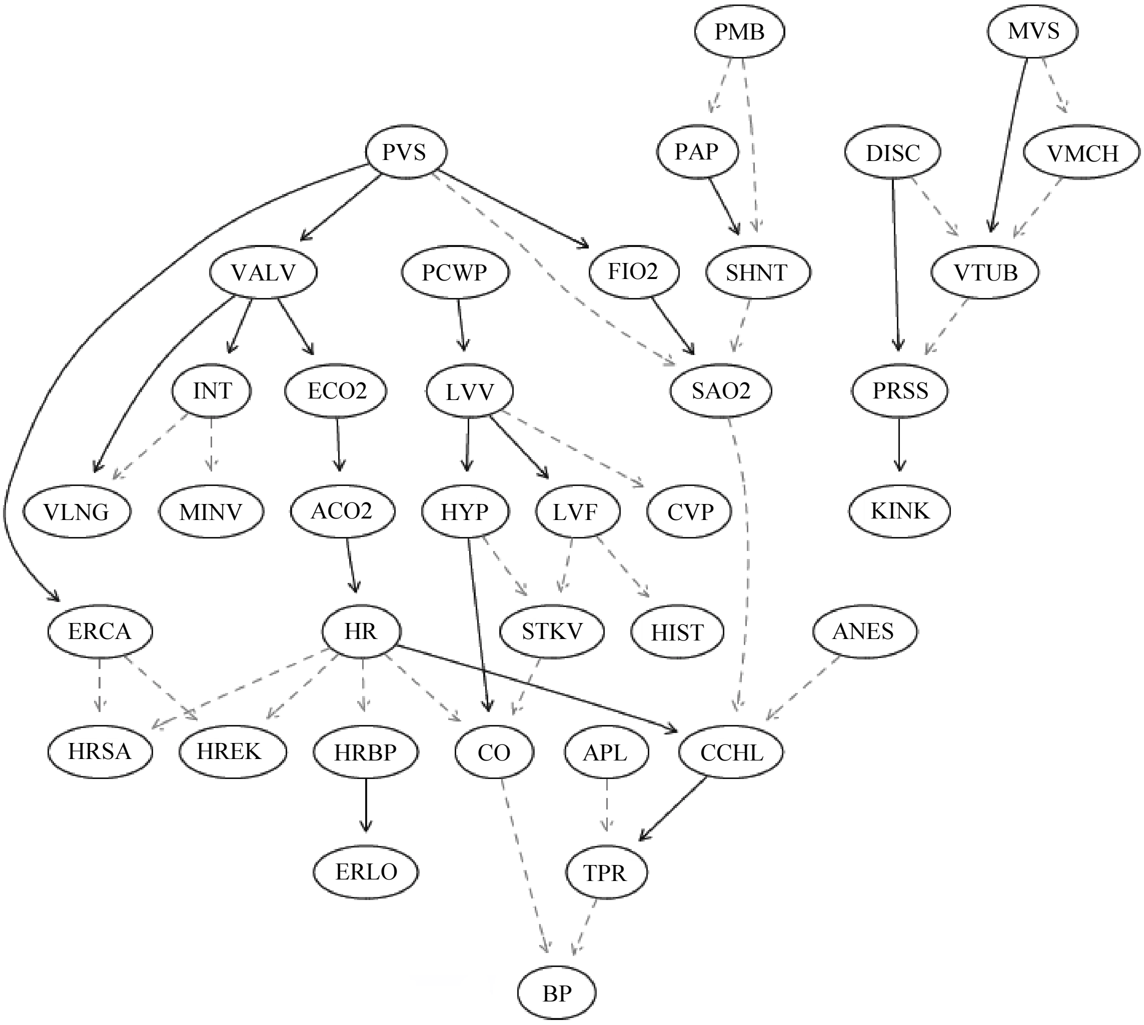


Fig. 7: MCMC-Fuzzy structure for the Alarm database

Table 3: Results for the Asia Network

Algorithm	Total	Correct arcs	Extra arcs	Missing arcs
K2	8	7	1	1
MMHC	6	4	2	4
EM-MCMC	8	4	4	4
MCMC-Fuzzy	8	6	2	2
Gold standard	8	8	0	0

Table 4: Results for the alarm network

Algorithm	Total	Correct arcs	Extra arcs	Missing arcs
K2	46	45	1	1
MMHC	35	18	17	28
EM-MCMC	36	19	17	27
MCMC-Fuzzy	46	26	20	20
Gold standard	46	46	0	0

The MCMC-Fuzzy algorithm yielded good results for both databases. When applied to the Asia database, it identified 6 correct arcs, which is compatible to the results for K2 with the advantage of not requiring additional parameters.

For the Alarm database, MCMC-Fuzzy found the largest number of correct arcs and the smallest number of missing arcs. However, these results are still inferior when compared to K2. MCMC-Fuzzy also identified more extra arcs than any other method.

Processing Time

The processing time tests were performed on a computer with an Intel Core i7 chipset and 8GB RAM memory, running MAC OS X 10.9. Table 5 shows the time in seconds per algorithm in each of the databases. EM-MCMC and MCMC-Fuzzy have the number of iterations as stop criteria. The number of iterations were 5,000 for the Asia database and 10,000 for the Alarm database.

Table 5: Processing time in seconds

Algorithms	Asia	Alarm
K2	1.5456	3.6432
MMHC	0.0385	1.2232
EM-MCMC	2.9568	15.1015
MCMC-Fuzzy	5.0924	27.7940

MMHC is known for optimal processing times and is widely used in large databases. By observing the processing time of the algorithms, one can note that the improvement in the structure had a cost, which is still acceptable. Since, structure learning only needs to be redone when there are changes in the data. In addition, the number of iteration for the MCMC methods influence directly upon their processing times.

K2 had the best results regarding processing times. However, the algorithm depends on previous knowledge over the database since the order or variables must be informed as an input parameter. The MCMC-Fuzzy method was proven efficient, yielding the best results in structure learning among the algorithms that learn Bayesian networks only from data. Furthermore, processing time could be improved by optimizations in the code.

Conclusions and Future Work

In this study, a fuzzy metric that combines distinct common metrics applied to Bayesian network structure learning is proposed. This metric was applied to a MCMC heuristic search algorithm, which is called Fuzzy-MCMC and evaluated using two synthetic databases.

The combination of different metrics resulted in a fuzzy system capable of evaluating different properties of Bayesian networks simultaneously. The modeled metric is also generic enough to be applied to other heuristic based search algorithms.

The proposed algorithm, Fuzzy-MCMC, has comparable results with the K2 algorithm for the Asia dataset. K2 has the best results for the Alarm dataset, with Fuzzy-MCMC having the second best results. However, the K2 algorithm requires as a parameter the order of the variables, which is often complex to determine. Considering algorithms that do not require additional parameters, the proposed MCMC-Fuzzy algorithm had the most accurate Bayesian network structure for both datasets. However, MCMC-Fuzzy algorithm also identified many extra arcs in comparison with other methods, hence, it obtained a more complex network structure. In relation to processing time, the MMHC algorithm had the best results. However, code optimizations can be performed to MCMC-Fuzzy in order to improve time performance. We believe that the MCMC-Fuzzy algorithm can be used when not much information is available about the dataset and when the extra processing time is not an issue.

Future work includes more experimentation with the fuzzy metrics' rules system, which can affect the accuracy of the structures learned as well as their running time. Future work also includes applying the proposed fuzzy metric to other heuristic search algorithms.

Acknowledgment

This research was supported by the Graduate Program of Computer Science (PPGCC) of the Federal University of Santa Catarina (UFSC).

Authors Contributions

Ademar Crotti Junior: Developed the method, participated in all experiments and contributed to writing of the manuscript.

Beatriz Wilges and Silvia Modesto Nassar: Participated in all experiments and contributed to writing of the manuscript.

Ethics

This article is original and contains unpublished materials. The corresponding author confirms that all other authors have read and approved the manuscript and there are no ethical issues involved.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control, 19: 716-723. DOI: 10.1109/TAC.1974.1100705

Aliferis, C.F., A. Statnikov, I. Tsamardinos, S. Mani and X.D. Koutsoukos, 2010. Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: Algorithms and empirical evaluation. J. Machine Learn. Res., 11: 171-234.

Alonso-Barba, J.I., L. de la Ossa, J.A. Gámez and J.M. Puerta, 2011. Scaling up the greedy equivalence search algorithm by constraining the search space of equivalence classes. Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Jun. 29-Jul. 1, Springer, Belfast, UK, pp: 194-205. DOI: 10.1007/978-3-642-22152-1_17

Bouckaert, R.R., 1993. Probabilistic network construction using the minimum description length principle. Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty, (ARU' 93), Springer, Berlin, Heidelberg, pp: 41-48. DOI: 10.1007/BFb0028180

Brooks, S., A. Gelman, G. Jones and X.L. Meng, 2011. Handbook of Markov Chain Monte Carlo. 1st Edn., CRC Press, ISBN-10: 1420079425, pp: 619.

Carvalho, J.V.D.F. and C. Chiann, 2013. Redes Bayesianas: Um método para avaliação de interdependência e contágio em séries temporais multivariadas. Rev. Brasileira Economia, 67: 201-217. DOI: 10.1590/S0034-71402013000200003

Castillo, E., J.M. Gutiérrez and A.S. Hadi, 2012. Expert Systems and Probabilistic Network Models. 1st Edn., Springer Science and Business Media, New York, ISBN-10: 1461222702, pp: 605.

Chib, S. and E. Greenberg, 1995. Understanding the metropolis-Hastings algorithm. Am. Stat., 49: 327-335. DOI: 10.2307/2684568

Chickering, D.M., 1996. Learning Bayesian Networks is NP-Complete. In: Learning from Data, Fisher, D. and H.J. Lenz (Eds.), Springer, New York, pp: 121-130.

Chickering, D.M., 2003. Optimal structure identification with greedy search. J. Mach. Learn. Res., 3: 507-554. DOI: 10.1162/153244303321897717

Daly, R., Q. Shen and S. Aitken, 2011. Learning Bayesian networks: approaches and issues. Knowl. Eng. Rev., 26: 99-157. DOI: 10.1017/S0269888910000251

Faulkner, E., 2007. K2GA: Heuristically guided evolution of Bayesian network structures from data. Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining, Mar. 1-Apr. 5, IEEE Xplore Press, Honolulu, HI, USA, pp: 18-25. DOI: 10.1109/CIDM.2007.368847

Friedman, N. and D. Koller, 2003. Being Bayesian about network structure. A Bayesian approach to structure discovery in Bayesian networks. Mach. Learn., 50: 95-125. DOI: 10.1023/A:1020249912095

Gámez, J.A., J.L. Mateo and J.M. Puerta, 2010. Learning Bayesian networks by hill climbing: Efficient methods based on progressive restriction of the neighborhood. Data Min. Knowl. Discovery, 22: 106-148. DOI: 10.1007/s10618-010-0178-6

Grzegorczyk, M. and D. Husmeier, 2008. Improving the structure MCMC sampler for Bayesian networks by introducing a new edge reversal move. Mach. Learn., 71: 265-265. DOI: 10.1007/s10994-008-5057-7

Guo, P. and N. Li, 2009. An EM-MCMC algorithm for Bayesian structure learning. Proceedings of the 2nd IEEE International Conference on Computer Science and Information Technology, Aug. 8-11, IEEE Xplore Press, Beijing, China, pp; 158-162. DOI: 10.1109/ICCSIT.2009.5234973

Heckerman, D., D. Geiger and D.M. Chickering, 1995. Learning Bayesian networks: The combination of knowledge and statistical data. Mach. Learn., 20: 197-243. DOI: 10.1023/A:1022623210503

Ko, S. and D.W. Kim, 2014. An efficient node ordering method using the conditional frequency for the K2 algorithm. Patt. Recog. Lett., 40: 80-87. DOI: 10.1016/j.patrec. 2013.12.021

Margaritis, D. and S. Thrun, 1999. Bayesian network induction via local neighborhoods.

Masegosa, A.S.R. and S. Moral, 2013. New skeleton-based approaches for Bayesian structure learning of Bayesian networks. Applied Soft Comput., 13: 1110-1120. DOI: 10.1016/j.asoc.2012.09.029

Morales, M.M., R.G. Dominguez and N.C. Ramirez, 2004. A method based on genetic algorithms and fuzzy logic to induce Bayesian networks. Proceedings of the 5th Mexican International Conference on Computer Science, Sept. 24-24, IEEE Xplore Press, Colima, Mexico, pp: 176-180. DOI: 10.1109/ENC.2004.1342603

Nielsen, J.D., T. Kočka and J.M. Peña, 2002. On local optima in learning Bayesian networks. Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence, Aug. 07-10, Morgan Kaufmann Publishers Inc., Acapulco, Mexico, pp: 435-442.

Niinimaki, T., P. Parviainen and M. Koivisto, 2012. Partial order MCMC for structure discovery in Bayesian networks. arXiv preprint arXiv:1202.3753.

Pellet, J.P. and A. Elisseeff, 2008. Using Markov blankets for causal structure learning. J. Mach. Learn. Res., 9: 1295-1342.

Scanagatta, M., C.P. de Campos, G. Corani and M. Zaffalon, 2015. Learning Bayesian networks with thousands of variables. Proceedings of the 28th International Conference on Neural Information Processing Systems, Dec. 07-12, MIT Press, Montreal, Canada, pp: 1864-1872.

Scutari, M. and J.B. Denis, 2014. Bayesian Networks: With Examples in R. 1st Edn., CRC Press, ISBN-10: 1482225581, pp: 241.

Sechidis, K. and G. Brown, 2015. Markov blanket discovery in positive-unlabelled and semi-supervised data. Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases, Sept. 07-11, Springer, Porto, Portugal, pp: 351-366. DOI: 10.1007/978-3-319-23528-8_22

Su, C. and M.E. Borsuk, 2016. Improving structure MCMC for Bayesian networks through Markov blanket resampling. J. Mach. Learn. Res., 17: 1-20.

Su, C., M.E. Borsuk, A. Andrew and M. Karagas, 2014. Incorporating prior expert knowledge in learning Bayesian networks from genetic epidemiological data. Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, May 21-24, IEEE Xplore Press, Honolulu, HI, USA, pp: 1-5. DOI: 10.1109/CIBCB.2014.6845507

Tsamardinos, I., L.E. Brown and C.F. Aliferis, 2006. The max-min hill-climbing Bayesian network structure learning algorithm. Mach. Learn., 65: 31-78. DOI: 10.1007/s10994-006-6889-7

Vafaee, F., 2014. Learning the structure of large-scale Bayesian networks using genetic algorithm. Proceedings of the Annual Conference on Genetic and Evolutionary Computation, Jul. 12-16, ACM, Vancouver, BC, Canada pp: 855-862. DOI: 10.1145/2576768.2598223

Xu, Z. and S.N. Srihari, 2014. Bayesian network structure learning using causality. Proceedings of the 22nd International Conference on Pattern Recognition, Aug. 24-28, IEEE Xplore Press, Stockholm, Sweden, pp: 3546-3551. DOI: 10.1109/ICPR.2014.610

Yan, L.J. and N. Cercone, 2010. Bayesian network modeling for evolutionary genetic structures. Comput. Math. Applic., 59: 2541-2551. DOI: 10.1016/j.camwa.2009.12.039

Yang, C.C., 1997. Fuzzy Bayesian inference. Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Oct. 12-15, IEEE Xplore Press, Orlando, Florida, USA, pp: 2707-2712. DOI: 10.1109/ICSMC.1997.635347

Yuan, C. and B. Malone, 2013. Learning optimal Bayesian networks: A shortest path perspective. J. Artificial Intell. Res. DOI: 10.1613/jair.4039

Zhang, Y., W. Zhang and Y. Xie, 2013. Improved heuristic equivalent search algorithm based on maximal information coefficient for Bayesian network structure learning. Neurocomputing, 117: 186-195. DOI: 10.1016/j.neucom.2013.02.015

Zhu, X. and Y. Yang, 2014. A fast Markov blanket discovery algorithm. Proceedings of the 5th IEEE International Conference on Software Engineering and Service Science, Jun. 27-29, IEEE Xplore Press, Beijing, China, pp: 318-322. DOI: 10.1109/ICSESS.2014.6933572