

# REFINING THE STRUCTURE OF BAYESIAN NETWORK LEARNT WITH AGGLOMERATIVE CLUSTERING TECHNIQUE FOR PROSTATE CANCER DISEASE

NAVEEN KUMAR BHIMAGAVNI<sup>1</sup>, ADILAKSHMI THONDEPU<sup>2</sup>

<sup>1</sup>Research Scholar, University College of Engineering, Osmania University, Department of Computer Science and Engineering, India

<sup>2</sup>Professor & HOD, Vasavi College of Engineering, Department of Computer Science and Engineering, India

E-mail: <sup>1</sup>naveenkumar0206@gmail.com, <sup>2</sup>t\_adilakshmi@gmail.com

## ABSTRACT

Structure learning of the Bayesian Network is a two-step process, one is parameter learning and the other is finding the best structure among search space using uncertain and incomplete data. Structure learning is the most important and complex task (*NP*-hard problem) in estimation theory. However, existing techniques require generating all possible graphs even for a small number of random variables, and consume a large amount of space and time complexity to verify each of them. Clustering-based Structure learning can be used to learn the structure of the Bayesian network to overcome this limitation. However, the learned structure needs to be refined as and when the new data arrives and existing refinement techniques verify the relation of a node with every other node; which consumes large time complexity.

In this work, we propose an algorithm that refines the structure of the Bayesian network learned using the agglomerative clustering technique using the proposed refinement algorithm. It considers only a subset of nodes (identified using Markov Assumption) for comparison of a node and thereby consumes comparatively less time complexity. Also, the Bayesian score is calculated for each candidate structure to find the best network structure.

**Keywords:** *Refinement Algorithm, Agglomerative Clustering, Bayesian network, Bayesian Score, Prostate Cancer, Markov Assumption*

## 1. INTRODUCTION

Bayesian network (also referred to as belief network), directed acyclic graphical model or hierarchical Bayes model, is a graphical model that encodes probabilistic relationships among random variables and their conditional dependence through a DAG (directed acyclic graph). Bayesian networks are constructed from the data or structure learned from a naive Bayes structure.

Domain Experts can construct Bayesian Networks manually using prior structure knowledge, especially in the medical domain. The structure can then be refined manually or automatically using techniques such as Refinement Algorithm and Expert Bayes [1]. However, manual construction may not be feasible for all domains. Another approach is to use available information. The structure of BN can also be estimated directly from the data. Most of the existing techniques [5][23]

The joint distribution of  $n$  binary-valued random variables will have  $2^n$  different assignments of values to  $n$  random variables and thereby the size of the distribution grows exponentially for large values of  $n$ . The goal is to reduce the size of the joint distribution by capturing *independent relations* among the random variables and the solution is a Bayesian network which is a compact representation of joint probability distribution.

**Definition:** Bayesian network (also referred to as the Bayesian belief network) [16] is a directed acyclic graphical (DAG) model that encodes conditional dependence among random variables. In this network, each node can represent a random variable and the edges represent causal relationships among these variables.

A  $n$ -dimensional Bayesian network (BN) is a triple  $B = (X, G, \Theta)$  where:

$X$  is a  $n$ -dimensional random vector where each random variable is ranged over by a finite domain.

$G$  is a directed acyclic graph (DAG) with  $n$  nodes and a set of edges  $E$  represents a set of casual dependencies among these variables and  $\Theta$  encodes the parameters of the network.

Advantages of Bayesian Network: Bayesian networks are used in various fields where the goal is to infer the values of uncertain variables. In addition, a Bayesian network is used in decision-making [26]. The major applications are medical diagnosis, financial management and risk assessment, weather forecasting, and sensor applications [24].

1. Missing data can be handled efficiently using as it encodes conditional independence relation among all the variables.

The Bayesian network is the best tool to capture casual relationships in a complex domain

Bayesian network is best suited for representing prior data and knowledge as it captures causal relationships among variables.

Bayesian networks and Bayesian statistical methods can be used to avoid over-fitting of the data.

Learning the structure of Bayesian is a complex task for applications involving a large number of random variables [3]. In such scenarios, instead of learning the structure from scratch, it is generally a good idea to utilize the knowledge of the domain experts. In cases, where the number of random variables is high, learning the structure is an  $NP$ -hard problem [25]. So, the Domain expert knowledge can be used to construct the initial structure; the initial structure can be further refined by existing specialized algorithms such as Expert Bayes [22]. The network structure created by domain experts needs to be refined for the following reasons [40].

1. New data might have been observed for some or all of the features (nodes) in the network and as a result, the existing structure may not be valid for the inference tasks [23].

2. The network structure created by experts might have missed some conditional dependence relationships among nodes or might have added unnecessary relationships which impact probabilistic inference even with complete data [25].

following are the steps to find the best structure. First, generate all Possible Graphs and then apply any scoring function for each structure [20][25]; The structure with the greater Bayesian score is considered the best Bayesian Network structure[22][26].

The task of structure learning for Bayesian networks refers to learning the structure of the directed acyclic graph (DAG) from observed data[2][4]. Bayesian networks are one of the powerful PGM tools for handling large and uncertain data[15][19]. There are three approaches for structure learning: score-based approach, constraint-based approach, and based on expert knowledge.[6][24].

The constraint-based case employs the independence test to identify a set of edge constraints for the graph and then finds the best DAG that satisfies the constraints [17][21]. For example, we could distinguish V-structure and fork-structure by doing an independence test for the two variables on the sides conditional on the variable in the middle [3]. This approach works well with some other prior (expert) knowledge of structure but requires lots of data samples to guarantee testing power[7][8]. So it is less reliable when the number of samples is small [14][18].

The score-based approach first defines a criterion to evaluate how well the Bayesian network fits the data and then searches over the space of DAGs for a structure with a maximal score[9][10]. In this way, the score-based approach is essentially a search problem and consists of two parts[11][12]: the definition of the score metric and the search algorithm[13][16]. Recent advances show that Bayesian network structures with a large number of variables can be learned with clustering techniques. However, clustering techniques learn the structure of the Bayesian network for each cluster separately and there is a need to refine the structure[36][40]. After learning the structure of the Bayesian network, it is required to refine the structure of BN as and when new data arrives [37][38].

There are some advantages of having an initial structure for a Bayesian network; for example, the structure is capable of capturing some meaningful correlations among variables [39][42] using the prior knowledge of the domain experts [41]; Refinement techniques can then be applied on the initial structure. As the initial structure is already created, searching for the best structure would be less costly[43][45].

Existing refinement techniques such as Expert Bayes uses the following technique to refine the initial network structure. It takes the initial network structure and data (training and test set). The conditional probability table for each node is updated with the case count frequency of the corresponding node [42][44].

For each combination of two nodes, it performs the following operations; add an edge in any direction if there is no edge; delete an edge if it already exists. Then check if a cycle is formed by performing these operations. For a node, update its conditional probability tables for each operation. if the Markov blanket [45] [49] of the node is updated. Compute the MLE (maximum likelihood score) [46][48] of the network structure after performing an operation. Select the network structure with the best score; repeat this process for every pair of nodes in the initial structure. Finally, apply the best network structure for the test set. In all cases, source and destination nodes are also chosen randomly. *bnlearn* package in R Programming Tool [29] [47] can be used to learn the structure of the Bayesian network. It is also used to learn the structure upon receiving the new data [32] [35]. Expert Bayes and *bnlearn* package randomly select two nodes and check if there exists an edge (relationship) between two nodes and performs edge deletion, add an edge, and reverse an edge direction. In a high dimensional space, verifying the relation of a node with the remaining nodes will be a higher computational task (leads to greater time complexity) [33][34]

### 1.1 Problem Statement

The main disadvantage of the score-based approach is that it consumes large time complexity. Agglomerative clustering-based algorithms can be used to overcome this problem and it consumes comparatively less time complexity. However, the learned structure needs to be refined as new data related to the domain is available. Existing refinement techniques verify the relation of a node with every other node in the Bayesian network and thereby consume large time complexity. In this work, a new refinement algorithm is proposed to refine the structure of a Bayesian network.

## 2. PROPOSED ALGORITHM

Initially, the structure of a Bayesian network is learned using the agglomerative clustering method.

Next, a refinement algorithm is applied to refine the structure of the Bayesian network.

The clustering algorithm has one limitation that, it only verifies the relationship among the nodes belonging to the same cluster. As a result, there is a possibility of missing dependency relationships with the nodes of different clusters. To capture such relations, there is a need for refinement of the initial network structure. So the refinement algorithm will re-verify the conditional independence relation among the nodes belonging to different clusters. To capture such missing dependencies, the clustering algorithm has been extended by adding a refinement process to the structure obtained by the clustering algorithm.

Step 1:

The proposed algorithm is implemented on the prevalent disease of Prostate cancer and gene expressions of the disease were collected from the data source

(<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128749>). The test data for Prostate cancer contains around nineteen thousand two hundred genes (19,200) and is stored in Google SQL cloud storage. Prepared test Gene expression Data for Prostate cancer disease in Google Cloud MySQL.

Step 2:

Find the ordering/ranking of the Genes based on the NormFinder Algorithm. It ranks the set of candidate normalization genes according to their expression stability in a given sample set and given experimental design. NormFinder can analyze expression data obtained through any quantitative method e.g. real-time RT-PCR and microarray-based expression analysis.

(<https://www.moma.dk/normfinder-software>)[6]

Step 3

Identify the clusters  $\{C_1, C_2, \dots, C_n\}$  among Prostate cancer Genome sequences using the Agglomerative clustering technique.

Step 4: for each cluster  $C_i$

begin

Find the highest ranking node among the cluster of nodes and mark it as the leader node  $L_i$

Find the structure of the Bayesian network for  $C_i$  using a proposed structure learning algorithm

Store structure of BN

end

Step 5:

<p>Find the structure of the Bayesian network for all the leader nodes <math>\{L_1, L_2, \dots, L_n\}</math></p> <p>Step 6: The refinement algorithm is applied to the learned structure</p> <p>Data: Bayesian Network structure learned using agglomerative clustering Newly available data</p> <p>output: Bayesian Network structure</p> <p>begin</p> <ol style="list-style-type: none"> <li>1. Read Bayesian Network structure (BNi);</li> <li>2. Read Newly available data (ND);</li> <li>3. Initialize Best_score (BS) = Compute Bayesian Score of the (Bni)</li> <li>4. Initialize Refined_Structure = (BNi)</li> <li>5. Compute sufficient statistics (Si) for all nodes using data (ND)</li> <li>6. for each node (Xi) in the network structure (BNi)</li> </ol> <p>begin</p> <ol style="list-style-type: none"> <li>7. generate a set of non-descendent nodes (Dk)</li> <li>8. for each non-descendant node (dk) in (Dk)</li> </ol>	<ol style="list-style-type: none"> <li>9. begin</li> <li>10. Add an edge from (Xi) to non-descendant node (dk) to form (New_BNi)</li> <li>11. Calculate Bayesian Score (Bi) using sufficient statistics (Si)</li> <li>12. if (Bayesian Score (Bi) &gt; Best_score (BS))</li> <li>13. Best_score (BS) = Bayesian Score (Mi)</li> <li>14. Refined_structure = (New_BNi)</li> <li>15. end</li> <li>16. end</li> <li>17. refined structure= BNi</li> </ol> <p>end</p> <p>In this step, the refinement algorithm will re-verify the conditional independence relation among the nodes belonging to different clusters. To capture such missing dependencies, the clustering algorithm has been extended by adding a refinement process to the structure obtained. For each node(Xi), a set of non-descendent nodes is computed using Markov Assumption. Add an edge from Node(Xi) to one of its non-descendent nodes and compute the Bayesian score of the new structure and compare it with the score of the original structure and mark the highest one as the current score. Repeat the same process by reversing the edge direction and computing the score.</p> <p>Similarly, repeat this process for all the nodes in the Bayesian network; finally, the structure with the highest score is marked as the best structure.</p> <p>Fig 1 displays the initial bayBayesiantnetwork with 20 nodes created by the clustering algorithm, after applying the refinement algorithm (RA) on it, the resultant refined Bayesian network structure has one extra edge added from the node AADACL2 (belongs to cluster 1) to node AIBG (belongs to the cluster 2) compare to the initial structure(Fig 1)</p>
--	---

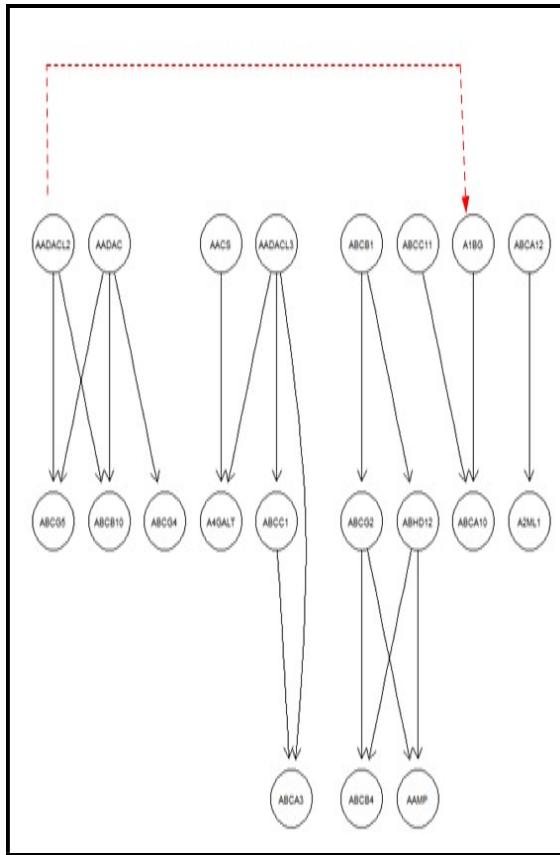


Figure 1: Refined structure: bayesian network structure with 20 nodes for Prostate cancer disease

### 3. RESULT ANALYSIS

The structure obtained by the proposed Refinement algorithm can be compared with the structure learned blearn algorithm and hill-climbing learned by the bnlearn package. Bayesian network models can be compared with the following metrics.

#### 3.1 Bayesian score

Bayesian score (3140.513) is computed for the 20-node Bayesian network created by the bnlearn package (Hill-climbing) and the clustering algorithm is displayed in the below table. note that, the cluster\_BN means it is the Bayesian network created by the proposed algorithm without refining the structure. Similarly, Refined\_cluster\_BN is the Bayesian network obtained after performing refinement. it is observed that score has been improved from 3486.921 to 3494.533 when the refinement algorithm is applied on the cluster\_BN.

The clustering algorithm (CA) has been extended to a greater number of gene expression nodes ranging from 20 to 100 and computed the score for each of the Bayesian networks. From the graph (Fig 2), it is observed that the gap between the set of score lines for bnlearn\_BN (Hill-climbing and tabu search). And cluster algorithm(Cluster\_BN and Refined\_cluster\_BN) line is getting increased linearly with the increasing number of nodes.

Table 1: Bayesian Score comparison table for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

Refined_cluster_BN	Cluster_BN	bnlearn_BN
3494.533	3486.921	3140.513

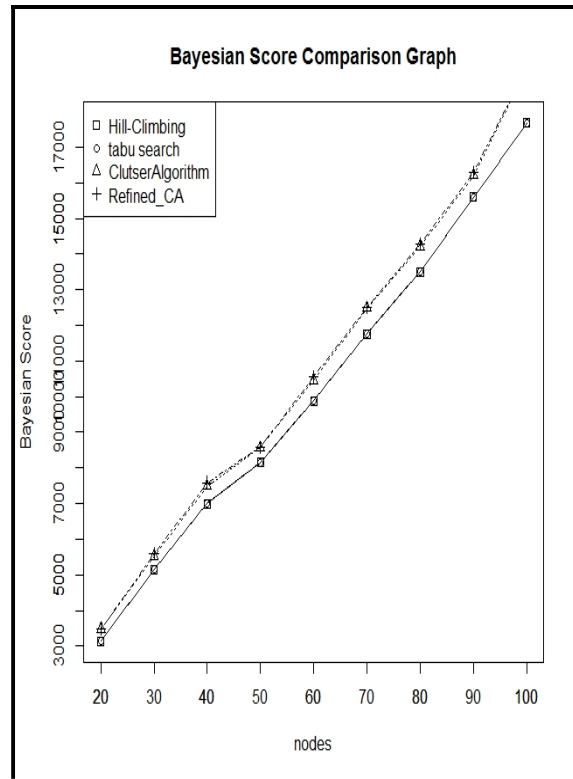


Figure 2: Bayesian sore comparison graph

#### 3.2 Expected loss

The expected loss is computed for the three Bayesian networks bnlearn\_BN, Cluster\_BN, and Refined\_cluster\_BN with 20 nodes and is displayed in the below Table. Cluster\_BN can reduce the expected error by 3% (from14.72803 to14.25463); further, when the refinement algorithm is applied to

the Cluster\_BN, the resultant Refined\_cluster\_BN can reduce the error by 3% (0.5341).

Table 3: Classification Error comparison table for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

Table 2: Expected loss comparison table for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

Refined_cluster_BN	Cluster_BN	Bnlearn_BN
13.72053	14.25463	14.72803

Refined_cluster_BN	Cluster_BN	Bnlearn_BN
0.1184493	0.1241949	0.1258183

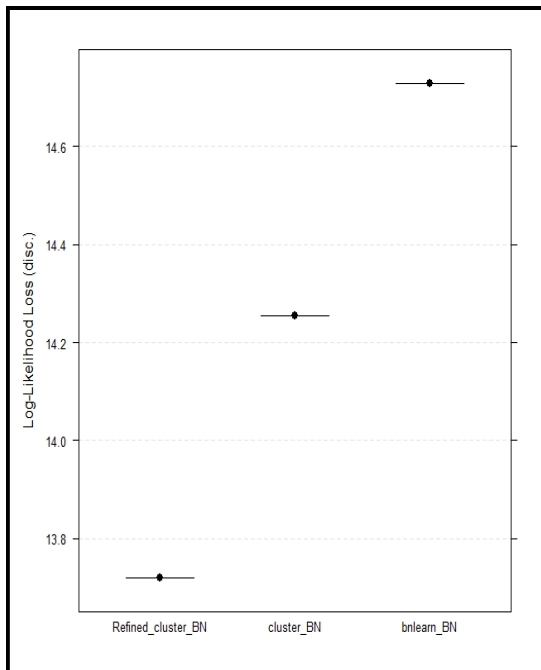


Figure 3: Comparison graph of Expected loss for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

From the comparison graph (Fig 3), it is observed that the refinement cluster Bayesian network can reduce the classification error compared to the cluster Bayesian network and bnlearn Bayesian network

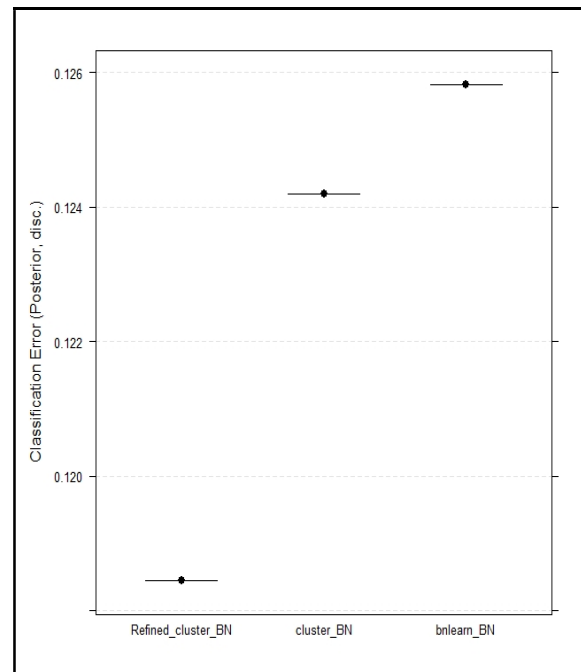


Figure 4: Comparison graph of classification error for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

### 3.3 Classification error

When classification error is computed for the target node *AIBG* in the refined clustering Bayesian network using the loss function (Classification Error(Posterior, disc)), the value is reduced by 5.6% (0.1184493 from 0.1258183). Classification error comparison table for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN.

### 3.4 Precision Analysis

Precision is a measure of how many predictions were correct out of the total predicted values using the Bayesian network model. For example, four by four confusion matrix (Table 3) is depicted for the target node (*AIBG*) using the refined cluster Bayesian network model with twenty nodes; in this matrix, the right diagonal elements represent *True*



positives for the column values; remaining values along the column represents *False positives*. For column 1 (0.40), true positives would be 165, and the rest of the values along the column (12,3 and 1) represents *False positives*.

Table 4: four x four confusion matrix for Refined\_cluster\_BN

OBSERVED	PREDICTED			
	0.40	1.22	2.85	4.48
0.40	165	0	0	0
1.22	12	3	1	0
2.85	3	1	2	1
4.48	1	0	3	3

Below (Table 5) displays precision values computed for the refined cluster Bayesian network, cluster Bayesian network, and bnlearn Bayesian network using Hill-climbing with 20 nodes.

Table 5: Precision comparison table for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

Refined_cluster_BN	Cluster_BN	Bnlearn_BN
0.911602	0.883720	0.881656

when the Refinement algorithm is applied on the clustered Bayesian network, an extra edge has been added from the higher rank node (AADACL2) to the lower rank node (AIBG); able to capture the missing dependency among the nodes. The same can be confirmed by computing the conditional probability tables for the node AADACL2.probablity value for the node (AADACL2= 0.40 given AIBG = 0.53) is improved from 0.75 to 0.86.

Also, it is observed that clustering a large number of nodes ( $n > 70$ ) gives better precision when compared to clustering a smaller number of nodes.

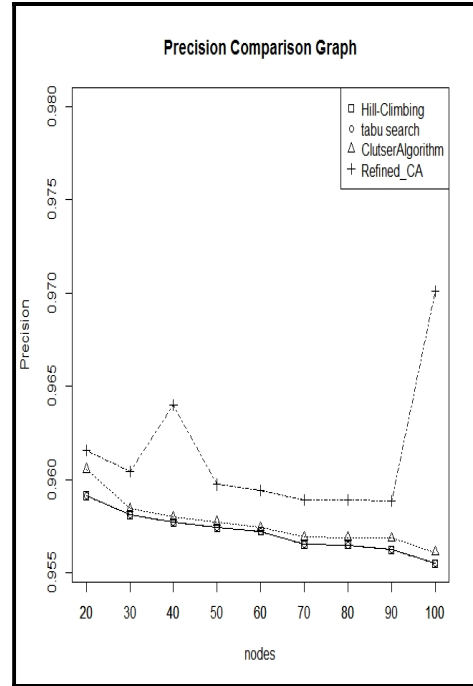


Figure 4: Precision comparison graph for different algorithms with the nodes ranging from  $n = 20$  to 100

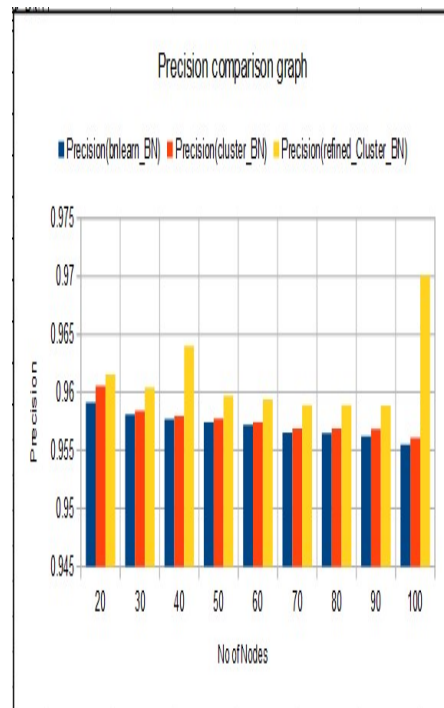


Figure 5: Precision comparison Graph for Refined\_cluster\_BN, Cluster\_BN, and bnlearn\_BN

4. CONCLUSION

In this work, Refinement Algorithm is used to refine the structure of the Bayesian network created by the agglomerative clustering technique with less complexity and it works on the principle of Markov Assumption. Also, the proposed algorithm ensures that minor updates occurred to the original network structure. It is assumed that the initial structure for the Prostate cancer Bayesian network is created by the agglomerative clustering technique. When the refinement algorithm is applied to this structure, one extra edge has been added to the original network (edge from *AADACL2* to *AIBG*).

The algorithm is capable of reducing the time complexity by  $O(n)$  performing an add edge operation for each node to one of its independent list nodes. Experiments have been carried out for the Prostate cancer Bayesian network and compare the goodness fit of the model with existing technologies like Expertbayes. An existing technique used the maximum likelihood score as a scoring function which tends to be overfitting the model. Our Refinement algorithm uses a Bayesian score as a scoring function to compute the score of each candidate structure. The score has been improved by 0.0028 (Table 3.3); though the improvement is quantitatively very small, this also ensures that minor changes are happening to the original structure. The algorithm further can be extended to hybrid Bayesian networks.

From the results, it is concluded that the Refinement algorithm can reduce expected loss and classification error when compared to existing techniques like Expertbayes, at the same time, precision and accuracy have been improved. Also, the refinement-based Bayesian network model can predict queries more accurately compare to the Expert Bayes. Inference [10] is another criterion to verify the model's performance. Refinement-based Bayesian networks can predict the target node probability values more precisely for both kinds of inferences (forward and backward propagation).

## REFERENCES:

- [1] TY - JOUR, AU - Almeida, Ezilda, AU - Ferreira, Pedro, T1 - Expert Bayes: Automatically Refining Manually Built Bayesian Networks VL - 2014 DO - 10.1109/ICMLA.2014.64 JO - Proceedings of the ... International Conference on Machine Learning and Applications. International Conference on Machine Learning and Applications.
- [2] UCI Machine Learning Repository: Arcene Cancer Data Sets <https://archive.ics.uci.edu/ml/machine-learning-databases/arcene/>
- [3] Abellán, J., Gómez-Olmedo, M., Moral, S.: Some variations on the PC algorithm. In: Third European Workshop on Probabilistic Graphical Models, pp. 1–8 (2006)
- [4] Adel, T., de Campos, C.P.: Learning Bayesian networks with incomplete data by augmentation. In: Proceedings of the 31st AAAI Conference on Artificial Intelligence, pp. 1684–1690 (2017)
- [5] Alonso, J., de la Ossa, L., Gámez, J., Puerta, J.: On the use of local search heuristics to improve GES-based Bayesian network learning. *Appl. Soft Comput.* 64, 366–376 (2018)
- [6] Alonso-Barba, J.I., de la Ossa, L., Puerta, J.M.: Structural learning of Bayesian networks using local algorithms based on the space of orderings. *Soft Comput.* 15(10), 1881–1895 (2011)
- [7] Andrews, D.F., Herzberg, A.: *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. Springer Series in Statistics, New York (1985)
- [8] Bacciu, D., Etchells, T., Lisboa, P., Whittaker, J.: Efficient identification of independence networks using mutual information. *Comput. Stat.* 28, 621–646 (2013)
- [9] Ben-Daya, M., Al-Fawzan, M.: A tabu search approach for the flow shop scheduling problem. *Eur. J. Oper. Res.* 109(1), 88–95 (1998)
- [10] Böttcher, S., Dethlefsen, C.: deal: A package for learning Bayesian networks. *J. Stat. Softw.* 8, 1–40 (2003)
- [11] Böttcher, S.: Learning Bayesian networks with mixed variables. In: Proceedings of the Eighth International Workshop in Artificial Intelligence and Statistics (2001)
- [12] Bottcher, S.G., Dethlefsen, C.: Deal: A package for learning Bayesian networks. *Journal of Statistical Software* 8, 200–3 (2003)
- [13] Burnside, E.S., Davis, J., Chhatwal, J., Alagoz, O., Lindstrom, M.J., Geller, B.M., Littenberg, B., Shaffer, K.A., Kahn, C.E., Page, C.D.: Probabilistic computer model developed from clinical data in a national mammography database format to classify mammographic



- findings. *Radiology* 251(3), 663–672 (Jun 2009)
- [14] C. Cotta and J. Muruzábal, “Towards a more Efficient evolutionary induction of Bayesian networks,” in *International Conference on Parallel Problem Solving from Nature*, pp. 730–739, Granada, Spain, 2002.
- [15] C. J. D’Orsi and L. W. Bassett and W. A. Berg et al.: *BI-RADS: Mammography*, 4th edition. American College of Radiology, Inc., 4th ed. (2003), Reston, VA 5. Chan, H., Darwiche, A.: Sensitivity analysis in Bayesian networks: From single to multiple parameters. In: *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*. pp. 67–75. UAI '04, AUAI Press, Arlington, Virginia, United States (2004), <http://dl.acm.org/citation.cfm?id=1036843.1036852>
- [16] C.-R. Hwang, “Simulated annealing: theory and applications,” *Acta Applicandae Mathematicae*, vol. 12, pp. 108–111, 1988.
- [17] C.-Y. Chan and Y. Ioannidis. Hierarchical cubes for range-sum queries. In *Proceedings of the 25th VLDB Conference*, pages 675–686, 1999.
- [18] Cheng, J., Bell, D.A., Liu, W.: An algorithm for Bayesian belief network construction from data. In: *Proceedings of Artificial Intelligence and Statistics*, pp. 83–90 (1997)
- [19] Chickering, D.: A transformational characterization of equivalent Bayesian network structures. In: *Proceedings of the Eleventh Annual Conference on Uncertainty in Artificial Intelligence*, pp. 87–98. Morgan Kaufmann (1995)
- [20] Chow, C., Liu, C.: Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theor.* 14(3), 462–467 (Sep 2006), <http://dx.doi.org/10.1109/TIT.1968.1054142>
- [21] Colombo, D., Maathuis, M.H.: Order-independent constraint-based causal structure learning. *Journal of Machine Learning Research* 15, 3741–3782 (2014)
- [22] Consortium, Elvira.: Elvira: An environment for creating and using probabilistic graphical models. In: Gámez, J., Salmerón, A. (eds) *Proceedings of the First European Workshop on Probabilistic Graphical Models*, pp. 222–230 (2002)
- [23] Continuous Learning of the Structure of Bayesian Networks: A Mapping Study By Luiz Antonio Pereira Silva, João Batista Nunes Bezerra, Mirko Barbosa Perkusich, Kyller Costa Gorgônio, Hyggo Oliveira de Almeida and Angelo Perkusich Submitted: April 26th 2018 Reviewed: July 9th 2018 Published: November 5th, 2018 DOI: 10.5772/intechopen.80064 Wai Lam and Fahiem Bacchus Department of Computer Science University of Waterloo Waterloo, Ontario, Canada, N2L 3G 1, Using New Data to Refine a Bayesian Network
- [24] Cooper, G.F., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* 9(4), 309–347 (1992), <http://dx.doi.org/10.1007/BF00994110>
- [25] Cooper, G.F.: The computational complexity of probabilistic inference using Bayesian belief networks. *Artif. Intell.* 42, 393–405 (1990)
- [26] Cussens, J., Malone, B., Yuan, C.: IJCAI 2013 tutorial on optimal algorithms for learning Bayesian networks (2013). <https://sites.google.com/site/ijcai2013bns/slides>. Accessed June 2018 de Campos, C.P., Corani, G., Scanagatta, M., Cuccu, M., Zaffalon, M.: Learning extended tree augmented naive structures. *Int. J. Approx. Reason.* 68, 153–163 (2015)
- [27] Cussens, J.: Bayesian network learning with cutting planes. In: *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, pp. 153–160 (2011)
- [28] Heckerman, D., Geiger, D., Chickering, D.: Learning Bayesian networks: the combination of knowledge and statistical data. *Mach. Learn.* 20, 197–243 (1995)
- [29] <https://moma.dk/files/Normfinder%20documentation-v20.pdf>; gene expression ordering of nodes ; <https://www.moma.dk/normfinder-software> : Norm finder algorithm
- [30] I. Tsamardinos, C. F. Aliferis, A. R. Statnikov, and E. Statnikov, “Algorithms for large scale Markov blanket discovery,” in *FLAIRS Conference*, vol. 2, pp. 376–380, St. Augustine, FL, USA, 2003.
- [31] J. Cheng, D. A. Bell, and W. Liu. An algorithm for Bayesian network construction from data. In *Artificial Intelligence and Statistics*, 1997. 105 106 Bibliography

- [32] J. M. Agosta. The structure of Bayes networks for visual recognition. In *Uncertainty in Artificial Intelligence*, pages 397–405. Elsevier Science Publishers B.V. (North-Holland), 1990.
- [33] Jaakkola, T., Sontag, D., Globerson, A., Meila, M.: Learning Bayesian network structure using LP relaxations. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, pp. 358–365 (2010)
- [34] Jaeger, M.: Probabilistic decision graphs—combining verification and ai techniques for probabilistic inference. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* 12, 19–42 (2004)
- [35] K. Beyer and R. Ramakrishnan. Bottom-up computation of sparse and Iceberg CUBEs. In *Proceedings of the 25th VLDB Conference*, pages 359–370, 1999.
- [36] K.-C. Chang and R. Fung. Symbolic probabilistic inference with continuous variables. In *Uncertainty in Artificial Intelligence (UAI)*, pages 77–85. Morgan Kaufmann, 1991.
- [37] K.-J. Kim, J.-O. Yoo, and S.-B. Cho, “Robust inference of Bayesian networks using speciated evolution and ensemble,” in *International Symposium on Methodologies for Intelligent Systems*, pp. 92–101, Saratoga Springs, NY, USA, 2005.
- [38] kaggle data source for covid19, <https://www.kaggle.com/iamhungundji/covid19-symptoms-checker>.
- [39] Kalisch, M., Bühlmann, P.: Estimating high-dimensional directed acyclic graphs with the PC algorithm. *J. Mach. Learn. Res.* 8, 613–636 (2007)
- [40] Koivisto, M., Sood, K.: Exact Bayesian structure discovery in Bayesian networks. *J. Mach. Learn. Res.* 5, 549–573 (2004)
- [41] Koivisto, M.: Parent assignment is hard for the MDL, AIC, and NML costs. In: *Proceedings of the 29th Annual Conference On Learning Theory*, vol. 4005, pp. 289–303 (2016)
- [42] Korhonen, J., Parviainen, P.: Exact learning of bounded treewidth Bayesian networks. In: *Artificial Intelligence and Statistics*, pp 370–378 (2013)
- [43] Kwisthout, J. H.P., Bodlaender, H.L., van der Gaag, L.C.: The necessity of bounded treewidth for efficient inference in Bayesian networks. In: *Proceedings of the 19th European Conference on Artificial Intelligence*, pp. 237–242 (2010)
- [44] Lauritzen, S., Wermuth, N.: Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Stat.* 17, 31–57 (1989)
- [45] Learning the Structure of Bayesian Networks: A Quantitative Assessment of the Effect of Different Algorithmic Schemes Stefano Beretta,<sup>1</sup> Mauro Castelli,<sup>2</sup> Ivo Gonçalves,<sup>3</sup> Roberto Henriques,<sup>2</sup> and Daniele Ramazzotti<sup>4</sup>.
- [46] Lee, C., van Beek, P.: Metaheuristics for score-and-search Bayesian network structure learning. In: *Proceedings of the 30th Canadian Conference on Artificial Intelligence*, pp. 129–141 (2017)
- [47] M. Drton and M. H. Maathuis, "Structure learning in graphical modeling," *Annual Review of Statistics and Its Application*, vol. 4, no. 1, pp. 365–393, 2017.
- [48] M. Gendreau and J.-Y. Potvin, Eds., *Handbook of Metaheuristics*, Springer, 2010.
- [49] M. Mascherini and F. M. Stefanini, “M-GA: a genetic algorithm to search for the best conditional Gaussian Bayesian network,” in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, pp. 61–67, Vienna, Austria, 2005.

