

An Efficient Coronary Heart Disease Prediction by Semi parametric Extended Dynamic Bayesian Network with Optimized Cut Points

K Gomathi¹, D Shanmuga Priyaa²

¹Department of Computer Science

Karpagam University, Karpagam Academy of Higher Education
Coimbatore, Tamil Nadu, India

²Department of Information Technology

Karpagam University, Karpagam Academy of Higher Education
Coimbatore, Tamil Nadu, India



ABSTRACT: Dynamic Bayesian Network (DBNs) is the general tool for enhancing the dependencies between the variables evolving in time and it's used to represent the complex stochastic processes to study their properties or make predictions based on the future behavior. The coronary heart disease (CHD) is considered as the one of the deadliest human diseases worldwide. The accurate prediction of CHD is very complex to be prevented and the treatment for it seems difficult. In early work, the TA methods with DBNs have been applied for the prognosis of the risk for coronary heart disease (CHD). The deviation of temporal abstractions from data is used for building DBN structure to predict CHD. However this approach cannot handle Complex temporal abstractions due to irregular time intervals. The cut-off values decided for temporal abstraction is the another issue in this work. In order to overcome this issue in this paper proposed the technique used for regularizing the irregular time interval in Extended Dynamic Bayesian Networks (DBNs) with temporal abstraction for coronary heart disease prediction. The proposed technique provides the global optimal solutions to assure the learning temporal solutions which provide observation of same irregularly spaced time points and the semi parametric subclass of the DBN proposed to allow further adaption of the irregular nature of the available data. The cut off value is searched from the domain expert knowledge base through the firefly optimization algorithm.

Keywords: Dynamic Bayesian Network, Firefly Optimization Algorithm, Temporal Abstraction

Received: 18 November 2019, Revised 5 February 2020, Accepted 1 March 2020

DOI: 10.6025/jes/2020/10/2/55-62

Copyright: with Authors

1. Introduction

The Bayesian network which represents the time through Dynamic Bayesian Networks (DBNs) [1] is used to connect the sequences of entire bayes networks. The each sequence is used to denote the snapshot in time. The temporal Abstraction technique is used to represent the time which is used to maintain the principles of Bayesian technique to analysis under uncertainly and delivers the simple network comparing to the DBNs.

The learning of DBN [2] is performed in two phases. Initially it refers the learning of static model and completes the identical manner with the traditional Bayesian Networks (BN's). Then the second phase observes the transition network which is used for temporal relationship between the random variables of different time slices.

In early work the TA methods with DBNs [3] have been applied for the prognosis of the risk for coronary heart disease (CHD). The deviation of temporal abstractions from data is used for building DBN structure to predict CHD. However this approach cannot handle Complex temporal abstractions due to its irregular time intervals. The cut-off values decided for temporal abstraction is the another issue in this work.

In order to overcome this issue, the global optimal solutions introduced to learning temporal systems which provide the fully observed at the same irregularly spaced time-points and a semi parametric subclass of DBN is introduced to allow further adaptation to the irregular nature of the available data. The cut-off value is searched from domain expert knowledge base through firefly optimization algorithm.

2. Literature Survey

Novel Approach to Predict Atherosclerosis [4] presents the new novel approach to predict the risk factors of the Atherosclerosis. The Coronary disease (CHD) is mainly caused by the Atherosclerosis disease. The prediction of risk factors reduces the death rate. The novel approach for finding the risk factors of the Atherosclerosis is built with the particle swarm optimization (PSO). In this technique the proposed methods are comparing with the machine learning techniques and these approaches are tested with STULONG dataset. Additionally the decision rules are integrated from the present technique to improve the accuracy further.

Dynamic Bayesian network to predict the cancer [5] presents the unique Pittsburgh cervical cancer screening model (PCCSM) to reduce the risk of the histopathologic cervical cancer. The PCCSM is the dynamic Bayesian network which is used to compute the quantitative cervical disease risk and estimate the risk components of the disease. The present technique is the graphic structure which provides the high risk HPV results, procedure data and histopathologic output. The present PCCSM technique generates the projection of negative risk and estimating the absence of histopathologic CIN2, CIN3 and adenocarcinoma.

Dynamic Bayesian Network based prediction of Coronary Atherosclerosis [6] proposed a novel approach for predicting the Atherosclerosis in the coronary arteries through the dynamic Bayesian networks. The main aim of the present technique is to investigate all the various sources of the information to predict the appearance of the disease. The present technique is used to identify the key factors to dictate the progression of the Atherosclerosis and based on the complex factors to find out the progression of the Atherosclerosis for the specific patient.

Predicting technique of cardiovascular disease [7] proposed a novel approach to predict the cardiovascular disease which is used to provide the advanced technique of possible to integrate biological systems on the global view then the combination with computational approaches also used to address the complexity of the disease. The present techniques provide the system based approaches to predicting the cardiovascular disease and evaluated the translational implications. The present technique improves the health care of the patients with the cardiovascular disease.

Dynamic Bayesian Network Based Disease Prediction [8] proposed new novel framework for medicine assisted diagnosis depends on the ontology and dynamic Bayesian network. The aims of the present techniques are divide the domain knowledge from probabilistic information and create the intuitive user interface. The present new framework has three layers such as the knowledge, uncertainty and user interface. The domain experts deliver the decoupled and the ontology constructs the domain concepts and captures only the domain knowledge. The Bayesian network is used the probabilities of the variable states and stored the profile repository. The result shows that the sample medicine model is used to diagnose the heart disease.

Prediction of Diagnose cardiogenic heart failure through Bayesian Network [9] proposed a novel approach for the prediction of the heart failure and to overcome from the computation imitation of the dynamic Bayesian network. The present technique contains both unobservable physiological variables and instrumentally observable events with the diagnosis like the myocardial infarction and future occurrence of the shock. The present technique shows that the prediction of occurrence of complex disease and perseverance of the heart disease are based on the variations of the feedback gathered from the patient.

Prediction of coronary artery disease [10] presents the prediction the high risk of the coronary artery disease (CAD). The coronary angiography performs the before pacemaker implantation. The CAD is described in the substantial proportion of the patients with the symptomatic brad arrhythmias and complex factors of the CAD. However these techniques are supported for the small dimensional dataset only.

Coronary Artery Disease Prediction [11] presents the systematic approach to predict the Coronary Artery Disease. The free text in the medical risk is used to identify the risk factors of the coronary heart disease automatically. The proposed ontology guide approach used to feature extraction and result is to compare with the traditional feature extraction techniques. The present techniques especially used in the feature extraction automatically filter the domain knowledge in the unified medical language system (UMLS). Since this technique has some limitations, such as which the classification it is difficult in the discriminate situations.

Dynamic Bayesian Network for the Prediction of Heart Disease [12] presents the dynamic Bayesian network for the prediction of the ischaemic heart disease (IHD). The main aim of the proposed technique is to investigate the potential association between the patient variables and survival of the risk of cardiovascular disease. The present techniques deliver the straight forward illustration of the casual relationships between the patients. The present techniques additionally provide the capacity to allow the representations of the repeated measured data and patients variables. However the proposed technique does not consider the potential improvement of the predictions.

Probabilistic Reasoning Methods for Prediction [13] presents the Probabilistic Reasoning Methods to predict the progression of the clinical findings (CF's) and the narrative portions of the medical records. The prediction based probability inference processing on the graphic model is used to encode the knowledge about the extraction of CF's from the PRs as well as the inferred chronological orderings, since the proposed technique does not reach the optimal solutions.

3. Proposed Methods

The proposed technique used for regularizing the irregular time interval in the Extended Dynamic Bayesian Networks (DBNs) with temporal abstraction for coronary heart disease prediction. The aim of the proposed methods is to develop the extended dynamic Bayesian network to regularize the irregular time interval and predicting the accurate cut of value.

The proposed methods includes the three main phases

- 1) Extended Dynamic Bayesian Network (EBN)
- 2) Semi parametric Extended Dynamic Bayesian network (SEBN)
- 3) Semi parametric Extended Dynamic Bayesian network with optimized cut points (SEBNOC)

3.1. Extended Dynamic Bayesian Network (EBN)

The temporal extension of the Bayesian network is the dynamic Bayesian network which denotes the process of the discrete time descriptions. The DBN structure is the repeated structure for the static Bayesian network for each time the certain interval. The DBN denotes the change of the states of the variables with different points. Each node in the DBN is called as the temporal process and it has the possible states.

Each node can be hidden in which the values are observed with the known values. The known values describe the local or classical dependencies within the variables. The DBN is assumed as the invariant to construct the network per time slice and the across time that does not change. Additionally the DBN uses the property of the Markovian process which is the conditional probability distribution of the each and every variables at time t for all the $t > 1$, depends on the parents from the same time slice or the previous time slice but not taken from the earlier time slices. The construction of the DBN consists of two phases such as the structure of the network and learning the parameter of the network.

3.1.1. Structure of the Network

The DBN framework enables to combine all the observations of the patients to predict the probability of the hypothesis of the patients and degrade the future CHD for all the given values of the observed nodes. The structure of the extended DBN network has 19 variables which constitute the nodes of the DBN. In this method the 17 are observed and 2 are hidden. The hidden variables of the class attributes are Pred_Event which denote the occurrence of the CHD events in the previous three years of the observations and the Dyslipidemia nodes. The above two variables are taken as the two values Such as the 'Present' and 'Absent'. In this method the Pred_Event variable is the terminal node and denotes the outsourced network with temporal and its only link with the parents in the last time slice of the unrolled network.

This method is used to denote the occurrence of the CHD event and it is repeated every time slice but the value is inferred at the

end of the inference process. The dyslipidemia delivers the new nodes to simplify the estimation process like the node of TCH, HDL, LDL and TRIG which is the indirect complex factors of the class attributes and current Event. The dyslipidemia is the cause for the *TCH, HDL, LDL, TRIG* and the variable of the class. Suppose the no confirmation of the one or more symptoms will used to increase the chances of the dyslipidemia which used to increase the increase the probability of the effects of the Current Event and Pred_Event.

3.1.2. Learning the Parameter of the Network

The structure of the network contains the conditional probabilities which quantify the arcs of the extended DBN network. The structure of the DBN is invariant for all times $t \in \{0 \dots 5\}$ the parameters of the network is allocated through each time slices and partitioned into two categories.

- 1) The prior probabilities for the root nodes
- 2) The transition probability for the root nodes.

Example: $Pr(Obesity_t | Obesit_{y-t}, DIET_t, Exerciset)$

The conditional probability distribution for the obesity based on the present status of the diet and exercise.

In this phase the each parameter is learned from the Expectation Maximization (EM). The expectation maximization contains two phases such as the expectation step, which is used to estimate the expected satisfaction over satisfactions of the missed data based on the values of the observed data then the maximization step, which is used to the parameters are estimated by the maximizing the expectation.

The network structures are described and quantified with the learned conditional probability distributions. The next step moves to find out the probability of the class node. *Pred_Event*. The each variable in the network initialized with the feature value. The DBN is unrolled for the time slices of six variable $t = [0, 1, \dots, 5]$. After that it performs the prognosis and derives the belief in the class variable Pred_Event which denotes the current value at $t = 6$. The Models are derived in the way of

$P(Pred_Event | Smoking_s, HT_s, Dyslipidemia_s, Obesity_s \dots)$

The probability of the CHD event delivers the confirmation of the last time slice.

3.2. Semi parametric Extended Dynamic Bayesian Network (SEBN)

The semi parametric extended dynamic Bayesian network is used to regularize the irregular time interval. The k time interval value used to calculate the support value. The support value is used to derive the exact interval time. The support values increase the time interval increased or the support values decreased means the interval values are also decreased.

3.3. Semi parametric Extended Dynamic Bayesian Network with Optimized Cut Points (SEBNOC)

In this semi parametric extended dynamic Bayesian network with optimized cut points (SEBNOC) is used to firefly algorithm to implement this model. The firefly algorithm is one of the new evolutionary algorithms ruled from predicting the coronary heart disease through the databases. This approach is called SEBNOC. The Firefly algorithm (FA) is inspired by the flashing pattern and behaviour of the fireflies. This algorithm considers the brightness of the firefly which is equal to the objective function value. The lower intensity firefly (lower fitness value) moves towards brighter firefly (higher fitness value). Here the each cut off values are assumed as the fireflies.

4. Proposed Algorithm

Step 1: Run the extended DBN network based on the regular time interval based on the K sized time interval.

Step 2: Initialize alpha (α) and attractiveness (β_0) and light absorption coefficient (γ) parameters. Initialize the rest of the cut off values as the random numbers.

Step 3: Find the fitness value for the each parameter.

The fitness value calculated based on the distance between the fireflies i and j which is calculated from

$$n_{i,j} = \sqrt{\sum_{1 \leq i \leq n} (x_i - y_i)^2}$$

Step 4: Randomly selects the parameters and records their fitness value, if there is the highest cut off value move towards the highest cut off value parameter (highest fitness value) based on the equation.

Here the X_i is the randomly selected cut off value and X_j is the highest cut off value.

$$\beta = \beta_0 e^{-\gamma_{i,j}}$$

$$X_{j,k}^h = (1 - \beta)X_{i,k}^h + \beta X_{j,k}^h + u_{j,k}^h$$

u -denotes the random numbers between 0 and $k=1, 2 \dots N$ and $h = 1, 2 \dots L$.

Step 5: If there is no firefly's fitness value is better than the selected firefly then it moves randomly in the search space based on the following equations.

$$X_{i,k}^h = X_{i,k}^h + u_{j,k}^h$$

$$k = 1, 2 \dots N_c \text{ and } h = 1, 2 \dots L$$

Step 6: Repeat step 3 to step 5 until the one of the termination criteria's is reached.

4. Result and Discussion

The proposed method Semi parametric Extended Dynamic Bayesian network (SEBN) and the Semi parametric Extended Dynamic Bayesian network with optimized cut points (SEBNOC) results are evaluated in terms of precision, recall, F1-score.

4.1. Dataset Description

The STULONG dataset is gathered from the longitudinal evaluation of atherosclerosis primary prevention. The target group associated to 1417 middle age men. The number of the check-up examinations of the each patients starts from the 1 to 20 age groups (1 to 24 years). The initial inspection of the every patient has measurements of the blood pressures, fundamental anthropometric measurements such as weight and height then the examination of the ECG.

Additionally the each patient delivers the information about the diet, physical activity, smoking, and alcohol drinking habits, social behaviour which may affect the job responsibilities as well as family and personal history. Furthermore the each value of the 244 attributes are surveyed in the initial investigations of each patient whereas on the following examinations, then the values of the 66 attributes are evaluated in the physical and biomedical investigations to find out the disease predictions.

4.2. Precision

Precision value is evaluated according to the feature classification of true positive prediction; false positive. It is expressed as follows:

$$\text{Precision} = \frac{\text{Truepositive}}{\text{Truepositive} + \text{Falsepositive}}$$

The figure 1 shows that comparison of precision values between the existing methods EBN and the proposed methods such as SEBN and SEBNOC, show that the proposed methods provides better results compare to existing methods in terms of precision values.

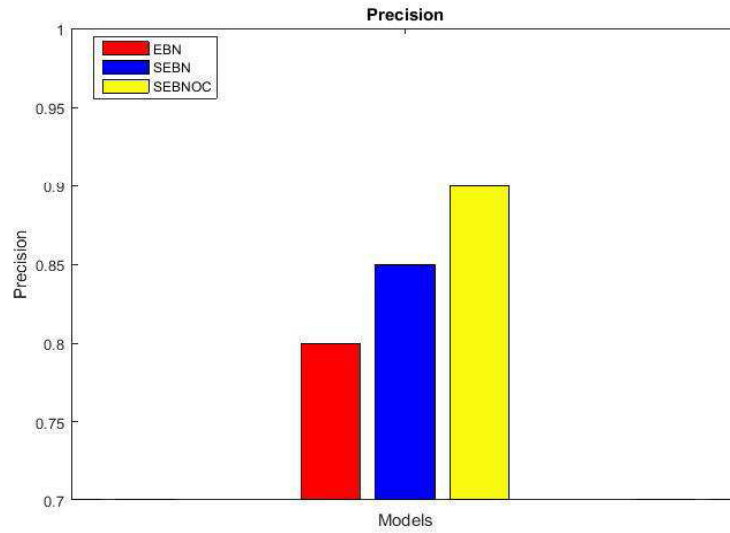


Figure 1. Comparison of Precision

4.3. Recall

Recall value is evaluated according to the feature classification of true positive prediction, false negative. It is given as,

$$Recall = \frac{Truepositive}{(Truepositive + Falsenegative)}$$

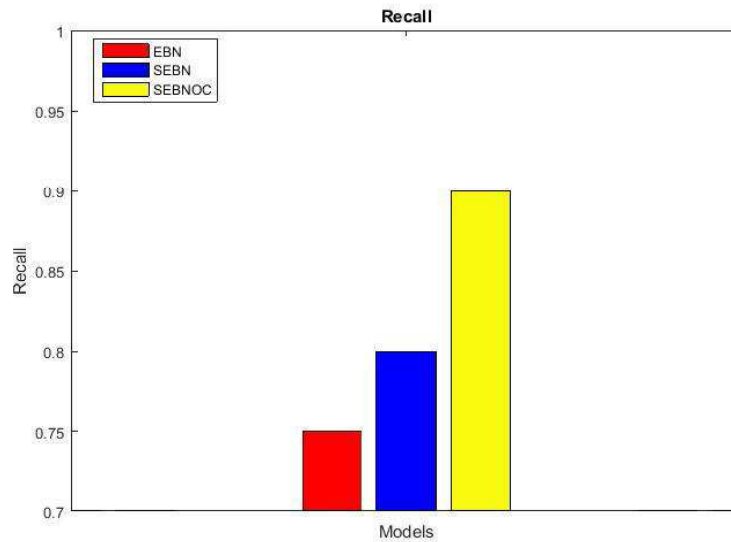


Figure 2. Comparison of Recall

The figure 2 shows that comparison of Recall values between the existing methods EBN and the proposed methods such as SEBN and SEBNO, show that the proposed methods provide better results comparing to existing methods in terms of Recall values.

4.4. F-Measure

F-measure is calculated from the precision and recall value. It is calculated as:

$$f - measure = 2 \times \left(\frac{precision \times recall}{precision + recall} \right)$$

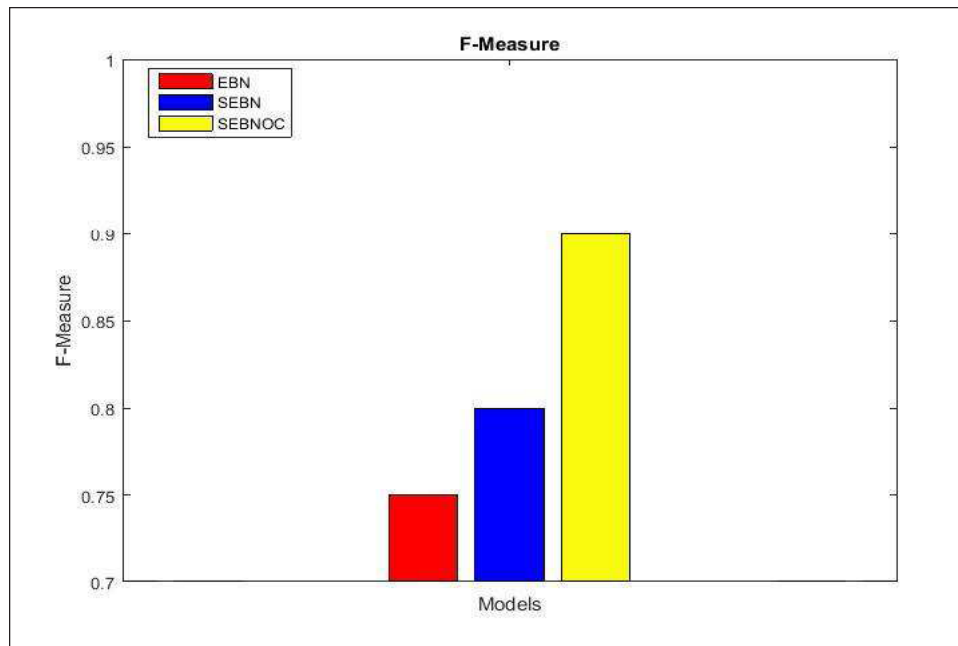


Figure 3. Comparison of F-Measure

The figure 3 shows that comparison of F-Measure values between the existing methods EBN and the proposed methods such as SEBN and SEBNOC, show that the proposed methods provide better results comparing to existing methods in terms of FMeasures values.

	Precision	Recall	F-Measure
EBN	0.80	0.74	0.75
SEBN	0.85	0.80	0.81
SEBNOC	0.91	0.92	0.90

Table 1. Shows the Comparison of EBN, SEBN, SEBNOC

5. Conclusion

The proposed Semi parametric Extended Dynamic Bayesian network (SEBN) and Semi parametric Extended Dynamic Bayesian network with optimized cut points (SEBNOC) overcome the limitation of the extended dynamic Bayesian networks to predict the coronary heart disease. The proposed technique used to regularise the irregular time interval and utilize the firefly algorithm to choose the cut off value in the networks. The experimental result shows that the proposed methods provide better results in terms of precision, recall and F-Measure.

References

[1] Grzegorzcyk, M., Husmeier, D. (2012). Bayesian regularization of nonhomogeneous dynamic Bayesian networks by

globally coupling interaction parameters. In *AISTATS* (pp. 467-476).

- [2] Robinson, J. W., Hartemink, A. J. (2010). Learning non-stationary dynamic Bayesian networks. *Journal of Machine Learning Research*, 11(Dec), 3647-3680.
- [3] Orphanou, K., Stassopoulou, A., Keravnou, E. (2016). DBN-extended: a dynamic Bayesian network model extended with temporal abstractions for coronary heart disease prognosis. *IEEE journal of biomedical and health informatics*, 20(3), 944-952.
- [4] Rao, V. S. H., Kumar, M. N. (2013). Novel approaches for predicting risk factors of atherosclerosis. *IEEE journal of biomedical and health informatics*, 17(1), 183-189.
- [5] Austin, R. M., Onisko, A., Druzdzal, M. J. (2010). The Pittsburgh cervical cancer screening model: a risk assessment tool. *Archives of pathology & laboratory medicine*, 134(5), 744-750.
- [6] Exarchos, K. P., Exarchos, T. P., Bourantas, C. V., Papafaklis, M. I., Naka, K. K., Michalis, L. K., Fotiadis, D. I. (2013, July). Prediction of coronary atherosclerosis progression using dynamic Bayesian networks. In Engineering in Medicine and Biology Society (EMBC), 2013, 35th Annual International Conference of the IEEE (p 3889-3892). *IEEE*.
- [7] MacLellan, W. R., Wang, Y., Lusis, A. J. (2012). Systems-based approaches to cardiovascular disease. *Nature Reviews Cardiology*, 9(3), 172-184.
- [8] Arsene, O., Dumitrache, I., Mihiu, I. (2011). Medicine expert system dynamic Bayesian Network and ontology based. *Expert Systems with Applications*, 38(12), 15253-15261.
- [9] Gatti, E., Luciani, D., Stella, F. (2012). A continuous time Bayesian network model for cardiogenic heart failure. *Flexible Services and Manufacturing Journal*, 24(4), 496-515.
- [10] Alai, M. S., Beig, J. R., Kumar, S., Yaqoob, I., Hafeez, I., Lone, A. A., Rather, H. A. (2016). Prevalence and characterization of coronary artery disease in patients with symptomatic bradyarrhythmias requiring pacemaker implantation. *Indian heart journal*, 68, S21-S25.
- [11] Buchan, K., Filannino, M., Uzuner, O. (2017). Automatic prediction of coronary artery disease from clinical narratives. *Journal of Biomedical Informatics*.
- [12] Marshall, A. H., Hill, L. A., Kee, F. (2010). Continuous Dynamic Bayesian networks for predicting survival of ischaemic heart disease patients. In Computer-Based Medical Systems (CBMS), 2010 *IEEE 23rd International Symposium on* (p 178-183). *IEEE*.
- [13] Goodwin, T., Harabagiu, S. M. (2015). A Probabilistic Reasoning Method for Predicting the Progression of Clinical Findings from Electronic Medical Records. *AMIA Summits on Translational Science Proceedings*, 2015, 61.