

CMDTS: The Causality-based Medical Diagnosis and Treatment System

Yaser Nemati^{✉1}, Pirooz Shamsinejad²

1) Department of Computer Engineering, Beyza Branch, Islamic Azad University, Beyza, Iran

2) Department of Computer Engineering and Information Technology, Shiraz University of Technology, Shiraz, Iran

nemati.y@gmail.com; p.shamsinejad@sutech.ac.ir

Received: 2017/04/09; Accepted: 2017/09/16

Abstract

Our medical world is replete with clinical data but this data is rarely automatically exploited for bringing more health to our society. Many researchers have been conducted in Medical Data Mining, but almost all of them have focused on diagnosing the diseases not treating the patients. In this paper we propose the Causality-based Medical Diagnosis and Treatment System, which can be used to diagnose a patient disease and suggest treatments to her/him. Our proposed system has three main subsystems: Causal Network Extractor, Diagnosis Subsystem and Treatment Suggesting Subsystem. Two main features of our system are: it takes solely observational data as input data and uses the causality-based action mining methodology. Action Mining is relatively a new trend in Data Mining which aims in proposing more actionable patterns to domain experts. We have implemented and tested our proposed method on some real and synthesized data. The results show superiority of our method over current state of the art method. Taking into account the causality results in more reliable treatments and makes it possible to use this system in real world situations.

Keywords: Medical Diagnosis System, Automatic Medical Treatment, Action Mining, Causal Networks

1. Introduction

Nowadays, we are in the age of data. The abundance of data which is stored without human ability to interpreting it makes repositories data tombs [1]. Data Mining (DM) emerged as a response to this need and became one of the dominant areas in computer science very soon.

Medical domain at the other hand showed attractions to DM from its first days of emergence. Diagnosis, Prognosis and Treatment are some medical sub-domains that DM and Machine Learning techniques have been used in them for decades. For example, authors in [2] have used SVM for detecting agitation transition in dementia patients. They fed three vital signs of patients- including Heart Rate, Galvanic Skin Response and Skin Temperature into a confidence-based SVM. In [3] supervised DM algorithms like Naive Bayes and K-NN have been used for heart disease diagnosis. Authors in [4] proposed a hybrid of Synthetic Minority Over-Sampling Technique (SMOTE) and Artificial Neural Network (ANN) to diagnose ovarian cancer from public available ovarian dataset. A Modified Fuzzy Logic method have been devised in [5] for

improving Breast Cancer Detection accuracy. Hosseinpour et. Al. [6] has been proposed a new method for diagnosing diabetes disease. However, literature is full of works in this domain, For some other ideas one can refer to these papers: [7-11].

While DM techniques tries to find unexpected and interesting patterns from huge amount of data, actionability of mined patterns is also desired, but traditional DM methods deliver descriptive patterns. So that, typically domain experts (Physicians in medical domain) need a lot of work to make decisions from these descriptive patterns. Action Mining (AM) came to existence as a response to this gap between descriptive mined patterns and actions that should be taken to gain a profit at target domain.

The first idea of AM proposed by Qiang Yang in [12] and Ras in [13]. Although there are some differences between these two works but the core idea is the same: Proposing some changes in the value of some attributes with the hope of gaining some profit in the corresponding domain. For example, in telecommunication churn prediction problem these actions can propose to change the service level of a customer from 'a' to 'b' to make it less probable that he/she will leave the company. Yang method tries to find most profitable actions for each user while Ras method (DEAR) tries to find action rules that are applicable to a range of users. After these two basic attempts for generating actions from traditional data some other researchers have been conducted in this domain.

Almost all of proposed methods in AM area deploy descriptive machine learning methods to find actions. For example, Yang uses Decision Trees for extracting actions [12], Ras et.al. uses classification rules[13-16]. Kalanat et.al. in [17], exploited the fuzzy decision tree to extract more applicable action rules. The most important shortcoming of using descriptive models to infer actions is to not consider causality. The reason is that descriptive models like Decision Tree and Classification Rules can only tell about correlation not about causation. When your task is prediction (normal task in DM) the difference between correlation and causation is not critical. Let's assume that we want to predict whether someone is sick or not. We will use his/her temperature as a powerful sign of illness. It means that temperature is strongly correlated with illness. But let's suppose that our task is to treat the patient (the goal in AM domain); Now, it doesn't make sense to try to decrease his body temperature to cure the illness, because temperature and illness have the same cause, i.e. the virus and by decreasing the temperature, you just push away a sign of illness not the illness.

Shamsinejad et. al. brought the idea of using causal relationship into AM world in [18-19]. In our research, Causal Networks as a probabilistic and casual graphical model has been used for extracting cost-effective actions. In this paper we extend their work to medical domain because causality is the most important factor in medical treatment. Our contribution is to devise a causal system for medical diagnose and treatment. The most unique feature of this system is that it is only based on non-experimental data. It means that we haven't any direct information about the effect of medical prescriptions on patients. We don't need to make some control experiments to investigate the effect of actions on patients; the system will find the most-profitable actions for patients. The rest of paper is as follows: in section 2 causal networks will be described, our medical diagnosis and treatment system will be presented in section 3, Experimental Results will be discussed in section 4 and Finally conclusion and future works come in section 5.

2. Causal Networks

A Bayesian Network (BN) is a graphical model that represents probabilistic relationships among variables of interest. It was first introduced by Pearl [20] and afterwards it is used in many domains including artificial intelligence, statistic, and philosophy. BN has the following properties that make it unique for knowledge representing among all other methods like SVM, Decision tree, ANN and other machine learning tools [21]:

1. It is well suited to deal with missing values in data that is a major problem in DM applications and specially in medical domain.
2. BN makes it possible to learn about causal relationships which are very important for understanding the problem domain.
3. By combining the BN with Bayesian statistical techniques we can integrate the prior domain knowledge with data and find more applicable patterns in real world through this.
4. BN in conjunction with Bayesian methods can be immunized against over fitting problem.

We suggest Casual Networks as the backbone for our medical diagnosis and treatment system. So that we should first explain the differences between a BN and Causal Network. Technically, BNs do not show the causal relationships. They are just a graphical and compact representation of joint probability distribution. In other words, a BN just states that if we know the values of parents of a node in BN, we can then compute the probabilities of different values of that node without any need to know the value of other nodes in BN. But Causal Networks are more restrictive than BN in this way that the edges in CNs must show the direct causation. So that a CN states that parents of a node are direct causes of it in the corresponding domain. It is clear that CNs convey more knowledge than BNs but creating them is also more complicated.

It was long deemed impossible to learn causal relationships from merely observational data, until Spirtes et al. [22] and Pearl [23] showed that under certain conditions one can infer some causal relationships from non-experimental data. These causal relationships are typically shown using a Directed Acyclic Graph (DAG) called a causal network (CN). In a CN, each node represents a variable and each edge represents a direct causal effect from the parent to the child. Causal networks are essentially Bayesian networks, with the additional guarantee that edges coincide with causal influences. Causal networks can be used like Bayesian networks for probabilistic inference. Queries for the (conditional) probabilities of events, prior and posterior marginal, most probable explanations and maximum a posteriori hypothesis can be answered (see [24] section 5.2 for more details). However, causal networks have the additional advantage that they can be used for reasoning about the effects of interventions: what can we infer if we set $A = a$, rather than observe $A = a$? Such interventions are basically the same as what we call actions in action mining. And in our medical system these type of actions can be used as treatments.

One of the most subtle tasks in causal inference is learning the structure of causal networks from merely observational data. Much work has been done in this area. The Inductive Causation (IC) algorithm proposed by Verma and Pearl [25] is the backbone of many proposed methods. This algorithm uses conditional independency relationships between attributes for finding causal links.

It is known that not all causal relationships can be discovered from observational data; causal relationships can be identified only under specific circumstances. As a result, IC returns only a partially directed acyclic graph (PDAG). In a PDAG, some edges are directed, and some undirected. Directed edges indicate causal relationships that have been identified with certainty. Undirected edges indicate that there is a relationship, but its type and direction cannot be determined with certainty. Our contribution is to use causal networks to extract medical actions. In the next section we present our causality-based medical diagnosis and treatment system and describe each module of the system in detail.

3. Causality-based Medical Diagnosis and Treatment System

Our proposed system is depicted in Figure 1. The system works as follows: First Subsystem, Causal Network Extractor (CNE), receives observational data from patient database. By observational we mean data that is not the result of controlled experiments but the traditional data that will be found in any medical center. CNE then extracts Causal Network from imported data. This causal network will be fed into other two subsystems: Treatment Suggesting Subsystem (TSS) and Diagnosis Subsystem (DS). DS will be used for prediction and diagnosis task. For example, to determine whether someone is sick or not. The TSS is designed for Treatment, where it uses causality-based action mining methodology devised by Shamsinejad et.al. [18] for suggesting treatment for patients. We will explain more elaborated each subsystem at the following.

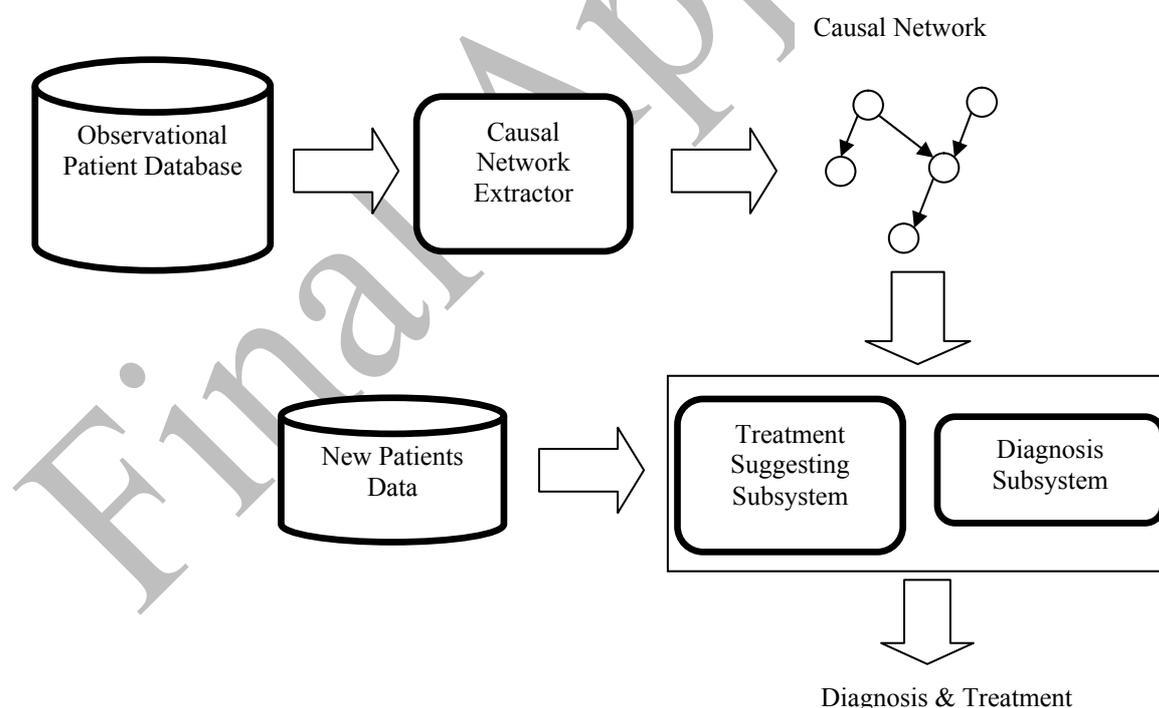


Figure1: Proposed Causality-based Medical Diagnosis and Treatment System

3.1 Causal Network Extractor

To extract causal network from observational data in CNE subsystem we use IC method by Verma and Pearl [23]. This method is outlined in Algorithm 1. The only

input to this algorithm is the probability distribution of system variables. It can be computed from observational data by counting the co-occurrence frequencies of variables in data samples. For more details about this computation one can refer to [18]. The result of CNE is a Causal Network that is the cornerstone of other two subsystems TSS and DS. In the following subsections We will explain how these subsystems use CN to do their jobs.

Algorithm 1: The IC Algorithm for extracting Causal Network from data

IC(P)
 Input: P , the Probability distribution of set of variables V .
 Output: A partially directed acyclic graph $H(P)$ compatible with P .
 Begin

1. Make a graph G with one node for each variable in V and no edges.
2. For each two variables a and b in V , search for a set S_{ab} in $V - \{a, b\}$ such that the conditional independence ($a \perp b \mid S_{ab}$) is held based on probability distribution P . It means a and b are independent given S_{ab} . Then draw an edge between a and b if there is no such set.
3. For each two non-adjacent node a and b with a common neighbor c , check if $c \in S_{ab}$:
 - If yes, continue.
 - Otherwise, set the direction of edges toward c . (i.e. $a \rightarrow c \leftarrow b$).
4. In the semi-directed graph resulted from the previous steps, make directed each undirected edge if both following conditions will be held:
 - No new V-structure will be created. (i.e. $a \rightarrow c \leftarrow b$)
 - No directed cycle will be created.

3.2 Diagnosis Subsystem

The mission of this subsystem in our proposed system is diagnosing the diseases of a patient. It receives CN and a patient as input and returns the most probable disease for that patient. It is clear that it will select one of the diseases that has been seen in the data and it cannot infer any unseen disease. Because the task of this subsystem is prediction there are no differences between Causal Network and Bayesian Network. It means that we can convert partially directed CN came from CNE module into one of the underlying BNs by making undirected edges directed in arbitrary directions and use that BN for inference about the most probable disease.

To Use the BN for diagnosis is much like Naïve Bayes classifier though with a big difference: In Naïve Bayes we assume that there is no correlation between non-class attributes. In other words, all attributes are directly correlated with class attributes. But in BN the correlation between non-class attributes also will be taken into account for prediction. Assume that the BN and its underlying probability distribution are at hand (it is the input from CNE module) and we want to diagnose a patient disease. The only thing that we need to do is computing the following formula:

$$Diagnosis_{BN}(P) = \arg \max_{d \in D} \{pr_{BN}(d | P)\} \quad (1)$$

Where P is the set of patient symptoms that DS wants to diagnose his/her disease. D is the set of all diseases in the data. pr_{BN} is the probability distribution computed in CNE and $pr_{BN}(d|P)$ is the conditional probability of seeing disease d given the evidence P (it is the probability that someone with symptoms P has the disease d). This probability can be computed easily by Most Probable Explanation (MPE) query in BN. Someone can refer to [24] for more details on computing this probability.

3.3 Treatment Suggesting Subsystem

One of our main contributions here is this part of the system. We use the Causality-based action mining methodology devised by Shamsinejad et.al. [18] for suggesting treatments to patients. Treatment is set of actions that should be taken by the patient. In this setting we can consider the treatment as action set in action mining terminology. So that we can use the ICE-CREAM method [18] for extracting treatments (cost-effective causality-based actions) from clinical data.

The core of ICE-CREAM is CREAM that we have adopted it for treatment learning with the name of Medical_CREAM in Algorithm 2.

Algorithm 2: Medical-CREAM for suggesting treatments for patients

```

Medical_CREAM (H, P, C, pg, CN)
Input: target health attribute H,
Patients data P,
cost data C : this is the cost of all possible actions. It can be fed to system by medical
experts if it is available. Although this is not mandatory for CREAM to work.
Health profit pg : the profit of treating a patient. We can set this to a constant.
underlying causal network CN : this is the input from CNE module.
Output: one treatment for each patient p ∈ P

1: for each p in P do
2:   I ← findCandidateActions(p, CN)
3:   Γ ← empty treatment
4:   finished ← false
5:   repeat
6:     αmax ← argmaxα ∈ I np(Γ ∪ {α}, p)
7:     if np(Γ ∪ {αmax}, p) > np(Γ, p) then
8:       Γ ← Γ ∪ {αmax}
9:       I ← I - {αmax}
10:    else
11:      finished ← true
12:    until I = ∅ or finished
13: assign treatment Γ to p

```

The task of `findCandidateActions()` in line 2 is to find all possible actions for a patient. For example, if one of the attributes in database is `TakeAcetaminophen` and its value for a patient is `false` (it means he/she hasn't taken Acetaminophen yet) then action (`TakeAcetaminophen, false → true`) is one of the possible actions. $np(\Gamma, p)$ is the net-profit of taking treatment Γ on patient p and is computed using the CN. For more detail about computing the net-profit refer to [18-19].

`Medical_CREAM` first finds all possible actions for a patient, then at each step add the most profitable action to the treatment. It will continue this until no improvement in net-profit of treatment happens (net-profit means profit minus costs). Assume we have a disease attribute in database that has 3 values: healthy, cold and flu. It is just enough to set the profit of healthy to Max Profit (for example 100), cold to 50 and flu to zero. In this setting our system will find treatments for patients to bring them from flu and cold state to healthy state.

4. Experimental Results

We have implemented and tested the proposed `Medical_CREAM` method on data and checked the actionability of resulted actions. Although for investigating the real effect of actions one should apply them on real world cases and check their results but here we have tried to synthesize data in a way that we can examine the proposed method. Our method to achieve that is as follows:

- Two type of causal networks have been considered: Real world causal networks like Chest Clinic, Headache and Alarm networks [26,27], and some synthesized causal networks in different sizes. Table1 shows the number of nodes and arcs for each network.
- For each causal network some records have been randomly sampled from the distribution defined by causal networks.
- The dataset which is created in last step have been fed into proposed system.
- For each causal network one node has been considered as target attribute and for each target attribute one value has been assigned the maximum profit.
- Because there is no cost data about current datasets we have generated cost matrices randomly for each dataset.
- For each dataset some objects have been considered as test cases and never fed to model. We then used them for evaluating our method.
- For evaluating our method, we have computed the net profit of mined actions using the true causal network.

Net profit of an action is the amount of profit we gain by applying the action minus cost we should pay for that action. Actions in medical domain are treatments and their cost are side effects. But, our different datasets will have different ranges of costs and profits. To make them more comparable, we express the effectiveness of an action set Γ on an object o using the Normalized Net Profit (nnp), which is defined as:

$$nnp(\Gamma, o) = \max\left(\frac{np(\Gamma, o)}{p_g}, 0\right) \quad (2)$$

Where np is the net profit of an action and will be computed by real causal network. For more details on how to compute nnp refer to [18].

Table1: Real and Synthesized Networks Specifications

Network Name	Number of Nodes	Number of Arcs
Chest Clinic	8	8
Headache	12	11
Alarm	37	46
Sample7	7	6
Sample15	15	27

We have compared our method with Yang method. Results have been shown in Table2. The results confirm that our method has outperformed Yang method in finding cost-profit actions. As it can be seen as much as the network gets more complex the average normalized net profit gets less for both methods which is expectable.

Table 1: Comparison of proposed method against Yang method based on Average Normalized Net Profit

Network	Medical_CREAM	Yang
Chest Clinic	0.58	0.41
Headache	0.72	0.22
Alarm	0.56	0.11
Sample7	0.35	0.25
Sample15	0.39	0.23

From the results it can be seen that the proposed method is successful in finding more cost-profitable treatments in comparison with Yang method. The reason is that Medical_CREAM based on causal relationships between attributes while Yang method only consider correlation for extracting actions.

5. Conclusion and Future Works

In this paper we proposed the Causality-based Medical Diagnosis and Treatment System with two main capabilities: Diagnosis and Treatment. It can predict the disease of the new patients in tandem with suggesting treatments for currently known patients. This system has the following unique features:

- 1- It is causality-based: It means that we take into account causal relationships between attributes. Because the system tries to find treatment it is very important to consider causality instead of correlation.
- 2- There is no need to predefined treatments or treatment database: One of the most important feature of this proposed system is that it can generate treatments solely from observational data. For example, when there is an attribute like TakeAcetaminophen in database it will create an action like (TakeAcetaminophen , false \rightarrow true) in candidate set. Then it can combine all of these candidate actions to find the most profitable treatment for a patient.
- 3- It is patient based: Our proposed system finds the best (i.e. the most cost-effective) treatment for each patient individually. This is much like the real life scenario when physician sees a patient.
- 4- It is cost aware: We can import the cost of actions into system and it will take into account those costs when suggesting treatment. For example, using this feature, we can inform the system that the cost of action (TakeAcetaminophen , false \rightarrow true) is much less than action (InjectDexamethasone , false \rightarrow true).

We have implemented our Medical_CREAM method for finding causal actions and tested on different causal networks. The results have shown that our method is more successful in finding cost-effective actions than current state of the art method –Yang method- in action mining domain.

For future works we will try to implement this system and make it applicable in real world. Some medical centers will be selected as pilot and the system will be tailored based on their needs. Besides, working on different theoretical aspects of the system is also under consideration for future works. For example, continuous attributes can be treated more intelligently. If we can integrate more Background Knowledge into the system, the results would be much more reliable. So that we will searching some ways to inject physicians background knowledge into our system more effectively.

References

- [1] J. Han, M. Kamber and J. Pei, “ Data Mining: Concepts and Techniques”, Morgan Kaufmann, 2012.
- [2] A. Rajkumar and G. S. Reena, “Diagnosis of heart disease using datamining algorithm,” Global Journal of Computer Science and Technology, vol.10, pp. 38-43, 2010.
- [3] M. Haitham, A. Angari and A. V. Sahakian, “Automated recognition of obstructive sleep apnea syndrome using support vector machine classifier,” IEEE Transactions On Information Technology In Biomedicine, vol. 16, pp. 463-468, 2012.
- [4] M. A. Hambali, M. D. Gbolagade, “Ovarian Cancer Classification Using Hybrid Synthetic Minority Over-Sampling Technique and Neural Network”, Journal of Advances in Computer Research, vol. 7, no. 4., pp 109-124, 2016.
- [5] F. Hoseini, A. Shahbahrani, A. Y. Notash and P. Bayat, “A Parallel Implementation of Modified Fuzzy Logic for Breast Cancer Detection”, Journal of Advances in Computer Research, vol. 7, no. 2, pp. 139-148, 2016.
- [6] N. Hosseinpour, M. Mosleh, S. Setayeshi, “Presenting a Real Time Method for Automatic Detection of Diabetes Based on Fuzzy Reward-Penalty System”, Journal of Advances in Computer Research, vol. 6, no. 3, pp. 125-135, 2015.
- [7] B. Kovalerchuk, E. Vityaev and J. F. Ruiz, "Consistent knowledge discovery in medical diagnosis," IEEE Engineering in Medicine and Biology Magazine, vol. 19, no. 4, pp. 26-37, 2000.
- [8] S. J. Fakhri and T. K. Das, "LEAD: a methodology for learning efficient approaches to medical diagnosis," IEEE Transactions on Information Technology in Biomedicine, vol. 10, no. 2, pp. 220-228, 2006.
- [9] M. X. Ribeiro, A. J. M. Traina, C. Traina Jr. and P. M. Azevedo-Marques, "An Association Rule-Based Method to Support Medical Image Diagnosis With Efficiency," IEEE Transactions on Multimedia, vol. 10, no. 2, pp. 277-285, 2008.
- [10] C. Tekin, O. Atan and M. Van Der Schaar, "Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis," IEEE Transactions on Emerging Topics in Computing, vol. 3, no. 2, pp. 220-234, 2015.
- [11] F. M. Khan and Q. Liu, "Transduction of Semi-supervised Regression Targets in Survival Analysis for Medical Prognosis," IEEE 11th International Conference on Data Mining Workshops, 2011.
- [12] Q. Yang, J. Yin, C. Ling and R. Pan, “Extracting actionable knowledge from decision trees,” IEEE Transactions on Knowledge and Data Engineering, vol. 19, pp. 43–56, 2007.

- [13] Z. Ras and L. Tsay, "Discovering extended action-rules (system dear)," *Intelligent Information Systems*, pp. 293–300, 2003.
- [14] Ras, Z.W. and A. Wieczorkowska. "Action-rules: How to increase profit of a company," *European Conference on Principles of Data Mining and Knowledge Discovery*, Springer, 2000.
- [15] Ras, Z.W. Ras, A. A. Tzacheva, L. S. Tsay and O. Girdal, "Mining for interesting action rules," *IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, 2005.
- [16] L. S. Tsay, and Z.W. Raś. "Action rules discovery system DEAR_3" *International Symposium on Methodologies for Intelligent Systems*, Springer, 2006.
- [17] N. Kalanat, P. Shamsinejadbabaki and M. Saraee, "A fuzzy method for discovering cost-effective actions from data", *Journal of Intelligent and Fuzzy Systems*, 2015.
- [18] P. Shamsinejadbabaki, M. Saraee, H. Blockeel, "Causality-based cost-effective action mining", *Intelligent Data Analysis*, Vol. 17, pp. 1075-1091, 2013.
- [19] P. Shamsinejadbabaki, "Mining Cost-effective Action Rules using Causal Structures", Ph.D. Thesis, Isfahan University of Technology, 2014.
- [20] J. Pearl, "Probabilistic reasoning in intelligent systems: networks of plausible inference," Morgan Kaufmann, 1988.
- [21] D. Heckerman, "A tutorial on learning with Bayesian networks", *Journal of Innovations in Bayesian Networks*, pp. 33-82, 2008.
- [22] P. Spirtes, C. Glymour, and R. Scheines, "Causation, prediction, and search," The MIT Press, 2000.
- [23] J. Pearl, "Causality: models, reasoning, and inference," Cambridge Univ Press, 2000.
- [24] A. Darwiche, "Modeling and reasoning with Bayesian networks," Cambridge University Press, 2009.
- [25] T. Verma and J. Pearl, "Equivalence and synthesis of causal models," *Uncertainty in artificial intelligence*, vol. 6, p. 255, 1991.
- [26] Hugin Lite 7.5 Samples, <http://www.hugin.com>
- [27] I. Suermondt and R. Chavez, The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks.